

STIC-Biot ch/Ch mLib

49689

From: Davis, Natalie  
Sent: Monday, August 27, 2001 2:29 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/589777

Please search SEQ ID NO:2

*Natalie A. Davis, PhD*

Patent Examiner

Art Unit 1642

CM1 Rm 8B13

ph (703) 308-6410

Point of Contact:  
Toby Port  
Technical Info. Specialist  
**CM1 1E01 TEL: 308-3534**

RECEIVED  
AUG 28 2001  
STIC

**THIS PAGE BLANK (USPTO)**

**CMU 1E01 TEL: 308-3234**  
**Technical Info. Specialist**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:49:51 : Search time 14.49 Seconds  
(without alignments)  
967.296 Million cell updates/sec

Title: US-09-589-777A-2

Perfect score: 968

Sequence: 1 HTTQDFQPVILVLAINTPLS.....CHNSYIVLCIENSFMTFSK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	968	100.0	1774	2 B56101	collagen alpha 1(X
2	965	99.7	1315	2 A56101	collagen alpha 1(X
3	840	86.8	684	2 A53019	collagen alpha 1(X
4	563	58.2	1388	2 A53317	collagen alpha 1(X
5	376	38.8	650	2 T22002	hypothetical prote
6	79	8.2	427	2 A53798	58k membrane-assoc
7	78	8.1	1055	2 S53597	chlorophyll a/b-bi
8	78	8.1	1955	1 ACCH	agrin precursor
9	77.5	8.0	474	2 T10011	probable penicilli
10	77.5	8.0	6420	2 T30283	polyketide synthas
11	77	8.0	208	2 T45272	methyltransferase
12	77	8.0	1575	2 S68448	synaptotagmin, 170K
13	76.5	7.9	1054	2 T30933	chitinase (EC 3.2.
14	76	7.9	309	1 E65112	hypothetical 34.6
15	76	7.9	309	2 B85985	hypothetical prote
16	75.5	7.8	183	4 S53318	hypothetical prote
17	75.5	7.8	4924	2 T50176	probable peptidase
18	75	7.7	587	1 SYHUA6	5-aminolevulinat
19	74.5	7.7	3670	2 T36249	CDA peptidase synth
20	74.5	7.7	7576	2 T17428	FK506 polyketide s
21	74	7.6	244	2 C45729	exopolysphatase
22	74	7.6	513	1 A45333	hypothetical prote
23	74	7.6	513	2 H85859	hypothetical prote
24	74	7.6	4180	2 G83559	pectinesterase hom
25	73	7.5	477	2 T05202	nitrile hydratase
26	72.5	7.5	419	2 D42725	gene CDS protein -
27	72.5	7.5	442	2 I47074	Xaa-His dipeptidase
28	72.5	7.5	463	2 B65994	nucleocapsid prote
29	72.5	7.5	739	1 VHWIEB	

30	72.5	7.5	800	2 T25140	hypothetical prote
31	72	7.4	732	2 C84487	hypothetical prote
32	72	7.4	2899	2 T21546	hypothetical prote
33	71.5	7.4	204	2 B45022	CRK-I - human
34	71.5	7.4	850	2 S56015	gastric mucin MUC5
35	71.5	7.4	5232	2 A45086	HC-toxin synthetas
36	71	7.3	512	2 A70569	probable cpsA prot
37	71	7.3	587	2 JX0278	5-aminolevulinat
38	71	7.3	646	2 D70939	hypothetical prote
39	71	7.3	841	2 B71212	hypothetical prote
40	70.5	7.3	1289	2 T18212	parasporeal crystal
41	70.5	7.3	252	2 F83350	hypothetical prote
42	70.5	7.3	257	2 S24812	minor fibrillar pro
43	70.5	7.3	316	1 F64966	probable transcript
44	70.5	7.3	319	2 B35090	MHC nonclassical c
45	70.5	7.3	416	2 G86065	probable cytochrom

## ALIGNMENTS

RESULT 1  
B56101  
collagen alpha 1(XVII) chain precursor, long splice form - mouse  
N:Contains: collagen alpha 1(XVII) chain precursor, medium splice form; endostatin  
C:Species: Mus musculus (house mouse)  
C>Date: 03-Oct-1995 #sequence revision 08-May-1998 #text-change 31-Mar-2000  
C:Accession: B56101; C56101; S72450; S65595; P0675; A54072; A58816  
J:Rein, M.; Pihlajaniemi, T.  
J: Biol. Chem. 270, 4705-4711, 1995  
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and ti  
lif homologous to rat and Drosophila fibrillized proteins.  
A:Reference number: A56101; MUID:95181468  
A:Accession: B56101  
A:Molecule type: mRNA  
A:Residues: 1-562 <REH1>  
A:Cross-references: GB:U11637; NID:9618429; PIDN:AAC52179.1; PID:9618430  
A:Experimental source: splice form clone PE17.24  
A:Accession: C56101  
A:Molecule type: mRNA  
A:Residues: 1-239,487-562 <REH2>  
A:Cross-references: GB:U11637; NID:9618429  
A:Experimental source: splice form clones PE8.1, PE19, PE15.2  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
submitted to the EMBL Data Library, August 1993  
A:Reference number: S72450  
A:Accession: S72450  
A:Molecule type: mRNA  
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524  
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-  
A:Reference number: A58370; MUID:94240111  
A:Accession: S65595  
A:Molecule type: mRNA  
A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>  
A:Cross-references: EMBL:L22545  
R:Abel, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Ninomiya, Y.  
Biochem. Biophys. Res. Commun. 196, 576-582, 1993  
A:Title: Identification of a novel collagen chain represented by extensive interrupt  
A:Reference number: P0675; MUID:94059075  
A:Accession: P0675  
A:Molecule type: mRNA  
A:Residues: 635-1774 <ABE>  
R:Rein, M.; Hintikka, E.; Pihlajaniemi, T.  
J. Biol. Chem. 269, 13929-13935, 1994  
A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial  
collagen chain.  
A:Reference number: A54072; MUID:94245707  
A:Accession: A54072  
A:Molecule type: DNA; mRNA  
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>

1

A:Cross-references: GB:003714; NID:9487733; PIDN:AAA20657.1; PID:9487734  
R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bl  
Cell 88, 277-285, 1997  
A>Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.  
A:Reference number: A56101; M0ID:97160848  
A:Accession: A56816  
A:Molecule type: protein  
A:Residues: 1591-1610 <ORE>  
A:Experimental source: hemangioidendothelium cells  
A:Note: Inhibits endothelial cell proliferation  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit o  
lated and subsequently O-glycosylated.  
C:Comment: The different splice forms of collagen alpha 1(XVII) may be involved in per  
C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of un  
may be useful in treating solid tumors.  
C:Genetics:  
A:Gene: MGI:Coll18a1  
A:Cross-references: MGI:71175  
A:Map position: 10:41.0  
A:Intons: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/  
A:Note: The list of introns is incomplete  
C:Superfamily: unassigned collagens  
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
F:1-1774/Product: collagen alpha 1(XVII) chain precursor, long splice form #status pred  
F:1-239,487-1774/Product: collagen alpha 1(XVII) chain precursor, medium splice form #s  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:361-486/Region: flizzled similarity  
F:786-812/Domain: collagenous #status predicted <CO01>  
F:823-896/Domain: collagenous #status predicted <CO02>  
F:921-1044/Domain: collagenous #status predicted <CO03>  
F:1066-1148/Domain: collagenous #status predicted <CO04>  
F:1163-1204/Domain: collagenous #status predicted <CO05>  
F:1218-1290/Domain: collagenous #status predicted <CO06>  
F:1301-1333/Domain: collagenous #status predicted <CO07>  
F:1346-1389/Domain: collagenous #status predicted <CO08>  
F:1351-1353/Region: cell attachment (R-G-D) motif  
F:1377-1428/Domain: collagenous #status predicted <CO09>  
F:1442-1459/Domain: collagenous #status predicted <CO10>  
F:1591-1774/Product: endostatin #status predicted <EST>  
F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity  
F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 968; DB 2; Length 1774;  
Best Local Similarity 100.0%; Pred. No. 3,5e-86;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHODFQPVHLVALNPLSGMKGIRGADFOCFQARAVGLSGTFRAPLSSRLQDLYSI 60  
|||||  
DB 1591 HTHODFQPVHLVALNPLSGMKGIRGADFOCFQARAVGLSGTFRAPLSSRLQDLYSI 1650  
QY 61 VRRADRGSVPIVNLKDEVLSPNSDLSFGSGQGLQPGARIFSPDGRDYLHPAMPQKSVW 120  
|||||  
DB 1651 VRRADRGSVPIVNLKDEVLSPNSDLSFGSGQGLQPGARIFSPDGRDYLHPAMPQKSVW 1710  
QY 121 HGSPSPRRRLMESYCFWRRTTGATGQASSLSGRLLGKRAKASHSYVLCLSENSFMT 180  
|||||  
DB 1711 HGSPSPRRRLMESYCFWRRTTGATGQASSLSGRLLGKRAKASHSYVLCLSENSFMT 1770  
QY 181 SFSK 184  
|||||  
DB 1771 SFSK 1774

RESULT 2  
A56101  
collagen alpha 1(XVII) chain precursor, short splice form - mouse  
N:contans: endostatin  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Oct-1995 #sequence, revision 08-May-1998 #text-change 31-Mar-2000  
C:Accession: A56101; A56371; S72450; S65595

R:Rehm, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995  
A>Title: Identification of three N-terminal ends of type XVII collagen chains and ti  
tif homologous to rat and Drosophila frizzled proteins.  
A:Reference number: A56101; M0ID:95181468  
A:Accession: A56101  
A:Molecule type: mRNA  
A:Residues: 1-103 <REH1>  
A:Cross-references: GB:011636; NID:618427; PIDN:AMC52178.1; PID:618428  
R:Rehm, M.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994  
A>Title: Alpha1(XVII), a collagen chain with frequent interruptions in the collageno  
A:Reference number: A56371; M0ID:94240112  
A:Accession: A58371  
A:Molecule type: mRNA  
A:Residues: 1-928 <REH2>  
A:Cross-references: GB:U1698; NID:9404754; PIDN:AAA37434.1; PID:9553894  
R:Oh, S.P.; Kamagata, Y.; Muraqaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
Submitted to the EMBL Data Library, August 1993  
A:Reference number: S72450  
A:Accession: S72450  
A:Molecule type: mRNA  
A:Residues: 28-687, 'U', 689-734, 'F', 736-751, 'R', 753-1315 <OH>  
A:Cross-references: EMBL:U22545; NID:9348968; PIDN:AAA19787.1; PID:9511298  
R:Oh, S.P.; Kamagata, Y.; Muraqaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A>Title: Isolation and sequencing of cDNAs for proteins with multiple domains of gly-  
A:Reference number: A58370; M0ID:94240111  
A:Accession: S65595  
A:Molecule type: mRNA  
A:Residues: 28-1315 <OH>  
A:Cross-references: EMBL:U22545  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
lated and subsequently O-glycosylated.  
C:Comment: The different splice forms of collagen alpha 1(XVII) may be involved in p  
C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of  
ay be useful in treating solid tumors.  
C:Genetics:  
A:Gene: MGI:Coll18a1  
A:Cross-references: MGI:71175  
A:Map position: 10:41.0  
C:Superfamily: unassigned collagens  
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:24-235/Region: thrombospondin amino-terminal similarity  
F:26-1315/Product: collagen alpha 1(XVII) chain, short splice form #status predicted  
F:327-353/Domain: collagenous #status predicted <CO1>  
F:364-437/Domain: collagenous #status predicted <CO2>  
F:462-583/Domain: collagenous #status predicted <CO3>  
F:607-689/Domain: collagenous #status predicted <CO4>  
F:704-745/Domain: collagenous #status predicted <CO5>  
F:759-831/Domain: collagenous #status predicted <CO6>  
F:842-874/Domain: collagenous #status predicted <CO7>  
F:887-910/Domain: collagenous #status predicted <CO8>  
F:982-894/Region: cell attachment (R-G-D) motif  
F:918-969/Domain: collagenous #status predicted <CO9>  
F:983-1000/Domain: collagenous #status predicted <CO10>  
F:1132-1315/Product: endostatin #status predicted <EST>  
F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity  
F:126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:172-228/Disulfide bonds: #status predicted  
F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 99.7%; Score 965; DB 2; Length 1315;  
Best Local Similarity 99.5%; Pred. No. 4.8e-86;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHODFQPVHLVALNPLSGMKGIRGADFOCFQARAVGLSGTFRAPLSSRLQDLYSI 60  
|||||  
DB 1132 HTHODFQPVHLVALNPLSGMKGIRGADFOCFQARAVGLSGTFRAPLSSRLQDLYSI 1191





A>Title: The agrin gene codes for a family of basal lamina proteins that differ in function  
 A:Reference number: A38857; MUID:92232298  
 A:Contents: alternative splicing  
 A:Accession: A38857  
 A:Molecule type: mRNA  
 A:Residues: 1132-1783; 1795-1955 <R02>  
 A:Cross-references: GB:M97371  
 A:Accession: B38857  
 A:Molecule type: mRNA  
 A:Residues: 1221-1647; 1652-1783; 1794-1955 <R03>  
 A:Cross-references: GB:M97372  
 A>Note: translation of the nucleotide sequence is not complete  
 R:Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.  
 Dev. Biol. 158, 523-535, 1993  
 A>Title: Developmental expression and alternative splicing of chick agrin RNA.  
 A:Reference number: 150692; MUID:93345745  
 A:Accession: 150692  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1784-1795 <THO>  
 A:Cross-references: EMBL:U07271; NID:q459665; PIDN:AA16788.1; PID:q459666  
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine receptors.  
 C:Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1 and agrin-related protein 2.  
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction  
 F:1-38/Domain: signal sequence #status predicted <SIG>  
 F:39-1783/Domain: agrin #status predicted <MAT>  
 F:39-1783/Domain: agrin #status predicted <MAT>  
 F:39-1647; 1652-1783/Domain: agrin-related protein 1 #status predicted <AG1>  
 F:77-126/Domain: kazal proteinase inhibitor homology <KPI1>  
 F:152-201/Domain: kazal proteinase inhibitor homology <KPI2>  
 F:223-273/Domain: kazal proteinase inhibitor homology <KPI3>  
 F:293-344/Domain: kazal proteinase inhibitor homology <KPI4>  
 F:343-483/Domain: kazal proteinase inhibitor homology <KPI5>  
 F:435-483/Domain: kazal proteinase inhibitor homology <KPI6>  
 F:500-548/Domain: kazal proteinase inhibitor homology <KPI7>  
 F:584-633/Domain: kazal proteinase inhibitor homology <KPI8>  
 F:673-726/Domain: laminin-type EGF-like homology <LE1>  
 F:729-773/Domain: laminin-type EGF-like homology <LE2>  
 F:801-851/Domain: kazal proteinase inhibitor homology <KPI9>  
 F:856-995/Region: serine/threonine-rich  
 F:1150-1219/Region: serine/threonine-rich  
 F:1233-1264/Domain: EGF homology <EG1>  
 F:1294-1448/Domain: laminin G repeat homology <LG1>  
 F:1429-1431/Region: motor neuron attachment (L-R-E) motif  
 F:1450-1482/Domain: EGF homology <EG2>  
 F:1489-1521/Domain: EGF homology <EG3>  
 F:1560-1711/Domain: laminin G repeat homology <LG2>  
 F:1718-1751/Domain: EGF homology <EG4>  
 F:1803-1955/Domain: laminin G repeat homology <LG3>  
 F:186-105; 94-126; 160-180; 169-201; 233-252; 241-273; 304-323; 312-344; 378-397; 386-418; 443-462; 1482; 1489-1500; 1494-1510; 1512-1521/Disulfide bonds: #status predicted  
 F:390; 659; 764; 814/Binding site: carbonylrate (Asn) (covalent) #status predicted

Query Match 8.1%; Score 78; DB 1; Length 1955;  
 Best Local Similarity 22.3%; Pred. No. 44;  
 Matches 47; Conservative 24; Mismatches 62; Indels 78; Gaps 9;

5 DEQPLHLVAVLTPISGGR--GIRGADQCQOARAGLSTFRF----- 49  
 DB 1670 DESKLAAALSTSTFYGVAVRISIKGVPLKEQHRSAVESIFRAHPCQKPNPCQNGC 1729  
 OY 50 -LSRLQDLYSIVRRADSGSVIVMLKDEVLSFSDSGSQ-----GQLQPGARIS 102  
 DB 1730 TCSPLLESTECACQK-----FSGHCKEVIETKAGDAEAIA 1767  
 OY 103 FDGR-----DYLRRHPAPQKSVWHGSDPSGRRLMESYCE--TWRTETTCATCQ 148  
 DB 1768 FDGRFYMEYHNAVTKSPALDPA-----EPSEKALQSNHPELSIKTEAT-----Q 1813  
 OY 149' ASLSLGRLEOKAASCHNSYIVLCIENSPM 179

DB 1814 GLILWCKGLR-----SDYIALAIWDGFV 1838  
 RESULT 9  
 T10011  
 probable penicillin-binding protein - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 R:Cole, S.T.  
 submitted to the EMBL Data Library, August 1997  
 A:Reference number: Z16916  
 A:Accession: T10011  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-474 <COL>  
 A:Cross-references: EMBL:Z70722; NID:e1059634; PID:e338505  
 C:Genetics:  
 A:Gene: pbpa

Query Match 8.0%; Score 77.5; DB 2; Length 474;  
 Best Local Similarity 24.0%; Pred. No. 8.8;  
 Matches 35; Conservative 18; Mismatches 66; Indels 25; Gaps 5;

OY 35 OQARVGLSTFRAPLFLSRDLSIVRRADSGSVIVMLKDEVLSFSDSGSQGOL 94  
 DB 101 EDAEDMLNGSDBRLFGRLADFT--GDPKRGAVDTTINPRVOTGMDAQGGGSSP 158  
 OY 95 QPCARIF-\*\*\*\*\*SFQGRDYLRRHPAPQKSVWH--GSDPS-----GRLMES 133  
 DB 159 CGAVVALEPSTGKILAWSTPSYDNLASHINPEQAOAMRLHDDPSPILNRAISFT 218  
 OY 134 Y--CEMRRETTGATGQASSLSGLR 157  
 DB 219 YPPGSTFKVITTTALQAGATTSDDL 244

RESULT 10  
 T30283  
 polyketide synthase - Streptomyces sp. (strain MA6548)  
 C:Species: Streptomyces sp.  
 A:Variety: strain MA6548  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Nov-2000  
 C:Accession: T30283  
 R:Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.  
 Eur. J. Biochem. 244, 74-80, 1997  
 A>Title: Structural organization of a multifunctional polyketide synthase involved in  
 A:Reference number: Z20806; MUID:97217427  
 A:Accession: T30283  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-6420 <MOT>  
 A:Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1  
 C:Genetics:  
 A>Note: fkbA  
 C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier pro  
 C:Keywords: carrier protein  
 F:51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>  
 F:1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
 F:3391-3462/Domain: acyl carrier protein homology <ACP1>  
 F:3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>  
 F:3983-4254/Domain: [acyl-carrier-protein]-S-malonyltransferase homology <AMT1>  
 F:5307-5378/Domain: acyl carrier protein homology <ACP2>  
 F:5431-5631/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>  
 F:5932-6206/Domain: [acyl-carrier-protein]-S-malonyltransferase homology <AMT2>  
 F:6293-6364/Domain: acyl carrier protein homology <ACP3>

Query Match 8.0%; Score 77.5; DB 2; Length 6420;  
 Best Local Similarity 23.0%; Pred. No. 2.1e+02;  
 Matches 47; Conservative 25; Mismatches 73; Indels 59; Gaps 9;

QY 4 QDEFVLHLVALNTPLSGGR-----GIRGADFOCGEQ-----A 37  
Db 5406 RQDDEPILAIYVAMACRLPGGVITSPPELRLVESGDTAITPAFGDRGMDLALYPPDPDAWG 5465  
QY 38 RAVGLSGTFRPAFLSSRLQDLXSYVRADRGSVYIVNLKDEVLSFSPSDLSFGSGOGLQPC 97  
Db 5466 KATSLGGFLEGAFAEDAFEDISPESLQMDPOQLR---LLETAMEAI---ERGRINPA 5519  
QY 98 ARIFSPGDRVYLHNPAPWPKSGRMLMESYCEFTWRETTGATGOASSLSGRL 157  
Db 5520 ----SLHGRI-----GVYGAAGSYGLGAE-----DREGNALTGTSLSLGR 5561  
QY 158 -----LEQKA-----ASCHNSIYVL 172  
Db 5562 AYVGLGEPAYVTDTACSSSLVAL 5585

RESULT 11  
T45272  
methylintransferase homolog [imported] - Streptomyces coelicolor (A3(2))  
C:Species: Streptomyces coelicolor  
A:Variety: A3(2)  
C:Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 11-May-2000  
C:Accession: T45272  
R:Martinez-Costa, O.H.; Martin-Triana, A.J.; Martinez, E.; Fernandez-Moreno, M.A.; Malpe  
J. Bacteriol. 181, 4353-4364, 1999  
A:Title: An additional regulatory gene for actinorhodin production in Streptomyces livida  
A:Reference number: 222953; MID:99328982  
A:Accession: T45272  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-208 <MAR>  
A:Cross-references: EMBL:Y18817; PIDN:CAB51133.1  
C:Genetics:  
A:Note: ONF8

Query Match 8.0%; Score 77; DB 2; Length 208;  
Best Local Similarity 25.6%; Pred. No. 3.6;  
Matches 40; Conservative 20; Mismatches 52; Indels 44; Gaps 7;  
QY 39 AVGLSGTFRALSSR-----LQDLYSIVRRADGSGV-----IVNLKDEVLSFSPSD---- 84  
Db 64 AVDISGVAERLAGHARTHGLDLDVAVRHDLRDSFEGHFDLVSA---HYLHPPFLDRA 121  
QY 85 SLFSGSGOGLQPCARI-----FSPGDRVYLHNPAPWPKSGRMLMESY 134  
Db 122 SYVGAHNLRLPGGRLLVVDHGSTAPMSWODDPDARHPAQEVAAADLADPA----- 173  
QY 135 CETWRE-----TTGATGOASSLSLGRLLQKA 162  
Db 174 --TWVRRAEAPRRRTATGPGGRTAEVVDHVLVRA 207

RESULT 12  
S68448  
synaptojanin, 170K - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Nov-1999  
C:Accession: S68448; S78547; S78527  
R:McPherson, P.S.; Garcia, E.P.; Slepnev, V.I.; David, C.; Zhang, X.; Grabs, D.; Sossin,  
Nature 379, 353-357, 1996  
A:Title: A presynaptic inositol-5-phosphatase.  
A:Reference number: A58183; MID:96149250  
A:Accession: S68448  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1575 <MCPE>  
A:Cross-references: EMBL:U45479; MID:g1166574; PIDN:AB60525.1; PID:g1166575; PID:g11665  
A:Experimental source: brain  
A:Accession: S78547  
A:Molecule type: protein

A:Residues: 1075-1093; 1173-1199; 1290-1308 <MCS>  
A:Experimental source: brain  
R:de Camilli, P.  
submitted to the EMBL Data Library, January 1996  
A:Reference number: S78527  
A:Accession: S78527  
A:Molecule type: mRNA  
A:Residues: 1-587; 589-1575 <DEC>  
A:Cross-references: EMBL:U45479  
C:Function:  
A:Description: binds the SH3 domain of amphiphysin  
C:Keywords: alternative splicing; alternative termination; synaptic vesicle  
F:1-1575/Product: synaptojanin, long form b #status predicted <MAT3>  
F:1-1139,1156-1575/Product: synaptojanin, long form a #status predicted <MAT2>  
F:1-1133,1156-1308/Product: synaptojanin, short form #status predicted <MAT1>

Query Match 8.0%; Score 77; DB 2; Length 1575;  
Best Local Similarity 22.6%; Pred. No. 43;  
Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;  
QY 8 PVLHLVALNTPLS--GSMRGIRGA-----DFOCFOQAR-AVGLSGTFRALSSRLQ 55  
Db 648 PFRDVAVDVTKTGCGATGKCAVAILRLPHRTSLDFVCSHFAAGSO-----VTERNE 702  
QY 56 DLXSYVRAD-----RGSVPVNLKDEVLSFSPSDLSFGSGO--- 91  
Db 703 DFEIARKLKSLFPMGRMLFSDHYFWCGDFMYRIDLPEVEVKELIRQGNMGLIAGQGLIN 762  
QY 92 ----GOLQPC-----ARIFS--FQGRDYLHNPAPWPKSVW 120  
Db 763 QKNAGOLFGRGLGKVFAPFYRYDLEFSEDDYDSEKCRTPAWTDVRLM 810

RESULT 13  
T30933  
chitinase (EC 3.2.1.14) A - Pseudomonas sp. (strain S9)  
C:Species: Pseudomonas sp.  
A:Variety: strain S9  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T30933  
R:Techakarnjanarak, S.; Goodman, A.E.  
Microbiology 145, 925-934, 1999  
A:Title: Multiple genes involved in chitin degradation from the marine bacterium Pseu  
A:Reference number: 220935; MID:99235578  
A:Accession: T30933  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1054 <TEC>  
A:Cross-references: EMBL:AF007894; MID:g9328771; PID:g9328775; PIDN:AAC79665.1  
C:Genetics:  
A:Gene: chIA  
C:Keywords: glycosidase; hydrolase

Query Match 7.9%; Score 76.5; DB 2; Length 1054;  
Best Local Similarity 24.6%; Pred. No. 29;  
Matches 41; Conservative 18; Mismatches 51; Indels 57; Gaps 8;  
QY 13 VALNTPLSGMRGIRGADFOCGFOQARV---GLSGTFRALSSRLQDLXSYVRADRG 68  
Db 206 VVLN--LSCSDQDQTVYSLFLFDGASVAQISGATGSGFNFTATSGNTQPSLATDQKA 263  
QY 69 VPVNLKDEVLSFSPSDLSFGSGOGLQPCARI-----FSPGDRV 108  
Db 264 T-----SDAQTLSPRIYDGTTPGDRDACKPEGLYQTPGVNTPYCTTIDADRED 313  
QY 109 L--RHAPWPKSVHNGSDPSGRRLMESYCEFTWRETTGATGOASSLSL 153  
Db 314 MGDHP-----RRVI-GYFTSMK---NGANGQPSYLV 341

RESULT 14





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:50:51 ; Search time 11.67 Seconds

(Without alignments) 540.104 Million cell updates/sec

Title: US-09-589-777A-2

Perfect score: 968

Sequence: 1 HTHODFQPVHLVLTPLS.....CHNSYLVLCIENSEFMTSFGK 184

Scoring table: BLOSUM62

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	968	100.0	1527	1	CAIH_MOUSE
2	840	86.8	1516	1	CAIH_HUMAN
3	563	58.2	1388	1	CAIE_HUMAN
4	78	8.1	1955	1	AGRI_CHICK
5	77	8.0	1574	1	SYJL_RAT
6	76.5	7.9	1050	1	ULK1_HUMAN
7	76	7.9	309	1	YHCC_ECOLI
8	75	7.9	1575	1	SYJL_HUMAN
9	75	7.7	505	1	CATL_METBA
10	75	7.7	587	1	HEMO_HUMAN
11	75	7.7	1324	1	SYJL_BOVIN
12	74.5	7.7	611	1	PHBC_RHIME
13	74	7.6	512	1	PPX_ECOLI
14	72.5	7.5	419	1	P47K_PSECL
15	72.5	7.5	739	1	VNUC_EBOG4
16	72.5	7.5	800	1	P5CS_GAEEL
17	72	7.4	1233	1	NME3_HUMAN
18	71.5	7.4	692	1	GTRB_BARBA
19	71.5	7.4	5217	1	HTSL_COCCA
20	71	7.3	587	1	HEMO_RAT
21	71	7.3	1289	1	CSAB_BACUD
22	70.5	7.3	257	1	FAEL_ECOLI
23	70.5	7.3	316	1	VEEY_ECOLI
24	70.5	7.3	319	1	HALL_RAT
25	69.5	7.2	437	1	ADFP_HUMAN
26	69.5	7.2	738	1	VNUC_EBOSB
27	69.5	7.2	739	1	VNUC_EBOZ5
28	69.5	7.2	739	1	VNUC_EBOZM
29	69.5	7.2	860	1	SZLA_BRARE
30	69	7.1	174	1	CARO_MYXA
31	69	7.1	692	1	VNUC_MABVM
32	69	7.1	695	1	VNUC_MABV
33	69	7.1	1023	1	DPOL_ADEB3

## ALIGNMENTS

RESULT	ID	CAIH_MOUSE	STANDARD	PRT	1527 AA
34	69	7.1	1357	1	VY03_YEAST
35	68.5	7.1	266	1	NANH_BACR
36	68.5	7.1	449	1	PURL_PYRO
37	68.5	7.1	1248	1	STJ2_RAT
38	68	7.0	362	1	SERC_BACCI
39	68	7.0	679	1	RNSA_MOUSE
40	68	7.0	1051	1	ULK1_MOUSE
41	68	7.0	1203	1	PPA2_YEAST
42	68	7.0	1220	1	C5AC_BACTU
43	68	7.0	1385	1	C5AA_BACUD
44	68	7.0	1443	1	SYJ2_HUMAN
45	67.5	7.0	230	1	VHEL_P1AMV

P47104 saccharomyc  
P31206 bacteroides  
O57979 pyrococcus  
O55207 rattus norv  
O59196 bacillus cl  
O05921 mus musculu  
O70405 mus musculu  
P22138 saccharomyc  
P36935 bacillus th  
O45760 bacillus th  
O15056 homo sapien  
O07517 plantago as





OY 181 S 181  
 DB 1514 A 1514  
 RESULT 3  
 CATE\_HUMAN STANDARD: PRT: 1388 AA.  
 ID CATE\_HUMAN  
 AC P39059;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE COLLAGEN ALPHA 1(XV) CHAIN PRECURSOR.  
 GN COL15A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=umbilical cord;  
 RC MEDLINE=94148920; PubMed=8106446;  
 RA Kivirikko S., Heleniak P., Rehn M.V., Honkanen N., Myers J.C.,  
 RA Pihlajaniemi T.;  
 RT "Primary structure of the alpha 1 chain of human type XV collagen and  
 RT exon-intron organization in the 3' region of the corresponding  
 RT gene";  
 RT J. Biol. Chem. 269:4773-4779(1994).  
 RN [2]  
 RP SEQUENCE OF 1-569 FROM N.A.  
 RC TISSUE=placenta;  
 RC MEDLINE=94140817; PubMed=8307960;  
 RA Murgaki Y., Abe N., Ninomiya Y., Olsen B.R., Ooshima A.;  
 RA "The human alpha 1(XV) collagen chain contains a large amino-terminal  
 RA non-triple helical domain with a tandem repeat structure and homology  
 RA to alpha 1(XVII) collagen";  
 RT J. Biol. Chem. 269:4042-4046(1994).  
 RN [3]  
 RP SEQUENCE OF 544-1252 FROM N.A.  
 RC MEDLINE=93066196; PubMed=1279671;  
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;  
 RT "Identification of a previously unknown human collagen chain, alpha  
 RT 1(XV), characterized by extensive interruptions in the triple-helical  
 RT region";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS  
 CC -1- SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: L25286; AAAS8429.1; -  
 CC EMBL: D21230; BAA04762.1; -  
 CC EMBL: L01697; -; NOT\_ANNOTATED\_CDS.  
 DR MIM: 120325; -  
 DR InterPro: IPR000087; -  
 DR Pfam: PF01391; Collagen: 4.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal.  
 FT CHAIN 1 25  
 FT SIGNAL 25  
 FT CHAIN 26 1388  
 FT DOMAIN 26 555  
 FT DOMAIN 556 573  
 FT DOMAIN 574 618  
 FT DOMAIN 574 618  
 FT NONHELIICAL REGION 1 (NC1).  
 FT TRIPLE-HELIICAL REGION 1 (COL1).  
 FT NONHELIICAL REGION 2 (NC2).

FT DOMAIN 619 732  
 FT DOMAIN 733 763  
 FT DOMAIN 764 798  
 FT DOMAIN 799 822  
 FT DOMAIN 823 867  
 FT DOMAIN 868 878  
 FT DOMAIN 879 949  
 FT DOMAIN 950 983  
 FT DOMAIN 984 1013  
 FT DOMAIN 1014 1027  
 FT DOMAIN 1028 1045  
 FT DOMAIN 1046 1052  
 FT DOMAIN 1053 1107  
 FT DOMAIN 1108 1117  
 FT DOMAIN 1118 1132  
 FT DOMAIN 1133 1388  
 FT DOMAIN 358 555  
 FT DOMAIN 358 555  
 FT REPEAT 358 408  
 FT REPEAT 409 459  
 FT REPEAT 460 509  
 FT REPEAT 510 555  
 FT REPEAT 306 306  
 FT CARBOHYD 324 324  
 FT CARBOHYD 687 687  
 FT CARBOHYD 807 807  
 FT CARBOHYD 814 814  
 FT CARBOHYD 1046 1046  
 FT CONFLICT 10 10  
 FT CONFLICT 49 49  
 FT CONFLICT 95 95  
 FT CONFLICT 150 150  
 FT CONFLICT 204 204  
 FT CONFLICT 409 409  
 SO SEQUENCE 1388 AA; 141930 MM; 60822AD925A3093D CRC64;  
 Query Match 58.2%; Score 563; DB 1; Length 1388;  
 Best Local Similarity 60.7%; Pred. No. 1e-45;  
 Matches 108; Conservative 23; Mismatches 43; Indels 4; Gaps 1;

OY 7 QYVLHVLAANTPLSGMRIRGADPQCFQARAVGLSCFFRFLSSRLDDLYSIYRRDR 66  
 DB 1215 KPALHLAALNMPFSGDIR---ADPCCFQARAAGLSTYRFLSLHDDLSTYRKAR 1270  
 OY 67 GSVPIYNLKDEVLSPEMSDILFSGSQCQLDQGARIFSFDSRDYLRIHPAMPQKSVHMGSDPS 126  
 DB 1271 YSLPIYNLKGVLFNMDISIFSGHGQPFNMHTPIYSFDSRDIMTDPSPQKXVHMGSSPH 1330  
 OY 127 GRLMESYCEWTEFTTGTATGQASSLLSRLLEOKRASCNSYIVLCIENSFTSFSK 184  
 DB 1331 GVALVNDYCEAWRTADTAVTGLASPLSTCKILDOKAYSCANRLIVLCIENSFTMDARK 1388  
 RESULT 4  
 AGRI\_CHICK STANDARD: PRT: 1955 AA.  
 ID AGRI\_CHICK  
 AC P31696;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE AGRIIN PRECURSOR.  
 GN AGRIIN.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=brain;  
 RC MEDLINE=92232297; PubMed=1314620;  
 RA Tsim K.W.K., Ruegg M.A., Escher G., Kroege S., McManan U.J.;  
 RT "cdna that encodes active agrin";

RL Neuron 8:677-689(1992).  
 RN (2)  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE-9223298; PubMed-1314621;  
 RA Ruesgg, M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,  
 RA Gensch E.M., McMahon U.J.;  
 RT "The agrin gene codes for a family of basal lamina proteins that  
 RT differ in function and distribution.";  
 RL Neuron 8:691-699(1992).  
 CC -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE  
 CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE  
 CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.  
 CC -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR  
 CC JUNCTION.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE DIFFERENT FORMS ARISE BY  
 CC ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR  
 CC CLUSTERING ACTIVITY.  
 CC -1- SIMILARITY: CONTAINS 8 KAZAL-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M94271; AAA48585.1; -;  
 DR EMBL; M97371; AAA48586.1; -;  
 DR EMBL; M97372; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; JH0591; AGCH.  
 DR InterPro: IPR000082; -;  
 DR InterPro: IPR000152; -;  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR001239; -;  
 DR InterPro: IPR001791; -;  
 DR InterPro: IPR002049; -;  
 DR InterPro: IPR002350; -;  
 DR Pfam; PF00008; EGF; 4.  
 DR Pfam; PF00390; SEA; 1.  
 DR Pfam; PF00050; Kazal; 9.  
 DR Pfam; PF00053; laminin\_EGF; 2.  
 DR Pfam; PF00054; laminin\_G; 3.  
 DR PRINTS; PR00290; KAZALINHBR.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_L; 6.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;  
 KW Laminin EGF-like domain.  
 FT SIGNAL 1 38  
 FT CHAIN 39 1955  
 FT DOMAIN 54 126  
 FT DOMAIN 130 201  
 FT DOMAIN 202 273  
 FT DOMAIN 276 344  
 FT DOMAIN 350 418  
 FT DOMAIN 419 483  
 FT DOMAIN 484 548  
 FT DOMAIN 551 633  
 FT DOMAIN 675 728  
 FT DOMAIN 729 775  
 FT DOMAIN 781 851  
 FT DOMAIN 856 995  
 FT DOMAIN 1150 1219  
 FT DOMAIN 1229 1265  
 FT DOMAIN 1446 1483  
 FT DOMAIN 1485 1522  
 FT DOMAIN 1714 1752  
 FT DISULFID 86 105  
 FT DISULFID 94 126  
 FT POTENTIAL.  
 FT AGRTN.  
 FT KAZAL-LIKE 1.  
 FT KAZAL-LIKE 2.  
 FT KAZAL-LIKE 3.  
 FT KAZAL-LIKE 4.  
 FT KAZAL-LIKE 5.  
 FT KAZAL-LIKE 6.  
 FT KAZAL-LIKE 7.  
 FT KAZAL-LIKE 8.  
 FT LAMININ\_EGF-LIKE 1.  
 FT LAMININ\_EGF-LIKE 2.  
 FT KAZAL-LIKE 9.  
 FT SER/THR-RICH.  
 FT SER/THR-RICH.  
 FT EGF-LIKE 1.  
 FT EGF-LIKE 2.  
 FT EGF-LIKE 3.  
 FT EGF-LIKE 4.  
 FT POTENTIAL.  
 FT POTENTIAL.

FT	DISULFID	160	180	POTENTIAL.
FT	DISULFID	169	201	POTENTIAL.
FT	DISULFID	233	252	POTENTIAL.
FT	DISULFID	241	273	POTENTIAL.
FT	DISULFID	304	323	POTENTIAL.
FT	DISULFID	312	344	POTENTIAL.
FT	DISULFID	378	397	POTENTIAL.
FT	DISULFID	386	418	POTENTIAL.
FT	DISULFID	443	462	POTENTIAL.
FT	DISULFID	451	483	POTENTIAL.
FT	DISULFID	507	527	POTENTIAL.
FT	DISULFID	516	548	POTENTIAL.
FT	DISULFID	592	612	POTENTIAL.
FT	DISULFID	601	633	POTENTIAL.
FT	DISULFID	675	687	BY SIMILARITY.
FT	DISULFID	677	694	BY SIMILARITY.
FT	DISULFID	696	705	BY SIMILARITY.
FT	DISULFID	708	726	BY SIMILARITY.
FT	DISULFID	729	741	BY SIMILARITY.
FT	DISULFID	731	748	BY SIMILARITY.
FT	DISULFID	750	759	BY SIMILARITY.
FT	DISULFID	762	773	BY SIMILARITY.
FT	DISULFID	810	830	POTENTIAL.
FT	DISULFID	819	851	POTENTIAL.
FT	DISULFID	1233	1244	BY SIMILARITY.
FT	DISULFID	1238	1253	BY SIMILARITY.
FT	DISULFID	1255	1264	BY SIMILARITY.
FT	DISULFID	1450	1461	BY SIMILARITY.
FT	DISULFID	1455	1471	BY SIMILARITY.
FT	DISULFID	1473	1482	BY SIMILARITY.
FT	DISULFID	1489	1500	BY SIMILARITY.
FT	DISULFID	1494	1510	BY SIMILARITY.
FT	DISULFID	1512	1521	BY SIMILARITY.
FT	DISULFID	1718	1731	BY SIMILARITY.
FT	DISULFID	1725	1740	BY SIMILARITY.
FT	DISULFID	1742	1751	BY SIMILARITY.
FT	CARBOHYD	390	390	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	659	659	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	764	764	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	814	814	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	VARSPLIC	1648	1651	MISSING (IN AGRIN-RELATED PROTEIN 2).
FT	VARSPLIC	1783	1793	MISSING (IN AGRIN-RELATED PROTEIN 1 AND AGRIN-RELATED PROTEIN 2).
FT	CONFLICT	1129	1131	RTI -> SII (IN AAA48586).
SO	SEQUENCE	1955 AA;	211411 MW;	B4DEB2/C23422581 CRC64;

Query Match 8.1%; Score 78; DB 1; Length 1955;  
 Best Local Similarity 22.3%; Pred. No. 22;  
 Matches 47; Conservative 24; Mismatches 62; Indels 78; Gaps 9;

QY	5	DFQVYLHVALNPLTSGMR--GIRGADFOCFQOARAVAGLSGTERAF-----	49
Db	1670	DESKLARAAMAAISTSEYGAORISIGVPLLEKQHIRSAVEISTFAHHCOTKPNPCQNGG	1729
QY	50	-LSSRLADLYSIVRADGSPYIVLKDDEVLSPSMDSLFSGSG-----GQLQPARIFIS	102
Db	1730	TCSPLLESTYECACORG-----FSGHCEKVIIEKAAGDAEIA	1767
QY	103	FDGR-----DYLRRPAPQKSVWHSQPSGRRLMESYCE-TWRTETTGATGQ	148
Db	1768	FDGRFYMEYHNNAVTKSPALDYPA-----EPSSKALQSNHFEISTENT-----Q	1813
QY	149	ASSLSGLRLEQKAKASCNSTIVLCIENSFPM	179
Db	1814	GLILMSGGLER-----SDYIALAIYDGFV	1838

RESULT 5  
 ID SYL\_RAT STANDARD; PRT: 1574 AA.  
 AC Q62910; Q62911; Q89092;  
 DT 30-MAY-2000 (Rel. 39, Created)

30-MAY-2000 (Rel. 39, last sequence update)  
 01-OCT-2000 (Rel. 40, last annotation update)  
 DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHOSPHATE 5-  
 DE PHOSPHATASE 1).  
 GN SYNJ.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;  
 RX MEDLINE=96149250; PubMed=8552192;  
 RA McPherson P.S., Garcia E.P., Slepnev V.I., David C., Zhang X.,  
 RA Grab D., Sosain W.S., Bauerfeind R., Nemoto Y., De Camilli P.,  
 RT "A presynaptic inositol-5-phosphatase.";  
 RL Nature 379:353-357(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=98374013; PubMed=9710239;  
 RA Moscholaki R., Finan P.M., Radley E., Parker P.J.;  
 RT "Identification and characterisation of a novel splice variant of  
 RT synaptojanin.";  
 RL FEBS Lett. 432:5-8(1998).  
 RN [3]  
 RP ALTERNATIVE SPLICING.  
 RC TISSUE=Brain;  
 RX MEDLINE=96394655; PubMed=8798761;  
 RA Ramjaun A.R., McPherson P.S.;  
 RT "Tissue-specific alternative splicing generates two synaptojanin  
 RT isoforms with differential membrane binding properties.";  
 RL J. Biol. Chem. 271:24856-24861(1996).  
 CC -1- FUNCTION: INOSITOL-5-PHOSPHATASE WHICH HAS A ROLE IN CLATHRIN-  
 CC MEDIATED ENDOCYTOSIS.  
 CC -1- CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O -  
 CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: LOCALIZED MAINLY IN THE SOLUBLE FRACTION (BY  
 CC SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 3 MAIN ISOFORMS; ISOFORM 1/170 KDA (SHOWN  
 CC HERE), ISOFORM 2/145 KDA AND ISOFORM 3/DELTA-SAC; ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. A STOP CODON IN POSITION 1309 IS SUPPRESSED  
 CC IN ISOFORM 1 DUE TO ALTERNATIVE SPLICING. A 16 AA INSERT IS  
 CC OPTIONALLY PRESENT IN THE THREE MAIN ISOFORMS.  
 CC -1- TISSUE SPECIFICITY: ISOFORM 1 IS FOUND IN NEONATAL BRAIN, AND IN A  
 CC WIDE VARIETY OF ADULT NON-NEURONAL TISSUES. ISOFORM 2 IS EXPRESSED  
 CC PREDOMINANTLY IN THE NEURONS, BUT IS ALSO FOUND IN ALL OTHER  
 CC TISSUES AT MUCH LOWER LEVELS. ISOFORMS 1 AND 2 ARE DETECTED IN THE  
 CC LUNG AND HEART. ISOFORM 1 IS EXPRESSED AT HIGHER LEVELS THAN  
 CC ISOFORM 2 IN THE TESTIS AND LIVER AND BOTH ISOFORMS ARE NOT  
 CC DETECTED IN THE SKELETAL MUSCLE. ISOFORM 3 WITH THE 16-AMINO-ACID  
 CC INSERT IS ONLY FOUND IN THE BRAIN WHILE ISOFORM 3 WITHOUT THE 16-  
 CC AMINO-ACID INSERT IS FOUND IN THE LUNG.  
 CC -1- DEVELOPMENTAL STAGE: AT EMBRYONIC DAY 12 (E12) ONLY ISOFORM 1 IS  
 CC SEEN WHILE AT E16 AND E18 ISOFORMS 1 AND 2 ARE SEEN. IN THE ADULT  
 CC BRAIN EXPRESSION OF ISOFORM 2 INCREASES DRAMATICALLY AS COMPARED  
 CC WITH ITS EXPRESSION IN EMBRYONIC BRAIN WHERE AS ISOFORM 1  
 CC DECREASES TO UNDETECTABLE LEVELS.  
 CC -1- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A  
 CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS (BY  
 CC SIMILARITY).  
 CC -1- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A  
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,  
 CC SH3P4, SH3P8, SH3P13 AND GRB2.  
 CC -1- DOMAIN: SPLICING OF THE SACI DOMAIN DOES NOT ALTER THE CATALYTIC  
 CC ACTIVITY OF SYNAPTOJANIN 1.  
 CC -1- SIMILARITY: IN THE CENTRAL SECTION, BELONGS TO THE INOSITOL-1,4,5-  
 CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 SACI DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U45479; AAB60525.1; -  
 CC EMBL: U45479; AAB60526.1; -  
 CC EMBL: AJ006855; CA07267.1; ALT\_TERM.  
 CC InterPro: IPR000300;  
 CC Pfam: PF00783; IPPC: 1.  
 CC DR PROSITE: PS50102; RRM; 1.  
 CC KW Hydrolyase; Alternative splicing; Repeat; Endocytosis; RNA-binding;  
 CC MultiGene family.  
 CC FT DOMAIN 1 499  
 CC FT DOMAIN 500 899  
 CC FT DOMAIN 894 971  
 CC FT DOMAIN 900 1574  
 CC FT DOMAIN 1033 1036  
 CC FT DOMAIN 1105 1110  
 CC FT DOMAIN 1123 1126  
 CC FT DOMAIN 1537 1545  
 CC FT DOMAIN 1401 1423  
 CC FT REPEAT 1401 1403  
 CC FT REPEAT 1410 1412  
 CC FT REPEAT 1421 1423  
 CC FT VARSPLIC 1 400  
 CC FT VARSPLIC 1140 1155  
 CC FT VARSPLIC 1309 1374  
 CC FT CONFLICT 588 588  
 CC FT CONFLICT 1140 1140  
 CC FT CONFLICT 1156 1156  
 CC SQ SEQUENCE 1574 AA; 172880 MW; 9DEESC876BAB3ADF CRC64;  
 Query Match 8.0%; Score 77; DB 1; Length 1574;  
 Best Local Similarity 22.6%; Pred. No. 22;  
 Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;  
 OY 8 PVLHVALNPNPLS--GGMGRIGRA-----DPCFOQAR-ANGLSGFFRAFLSSRLQ 55  
 DB 648 PSTRDAVDIVKGTGMGATGKGAVALRMLEHTTSLCFVCSHFAAQSQ-----VERNE 702  
 OY 56 DLSTIVRRAD-----RGSVPIVNLKDEVLSPMSDLFSGSQ--- 91  
 DB 703 DPEFIARKLSFPMGRMLFSHDYFVNGCDPFRYRIDLNEEKEKLIRGONDSLALAGQLIN 762  
 OY 92 ----GQLQPG-----ARIFS--FDGRDYLRRHPAMPQKSVW 120  
 DB 763 QKNAGQIFRCFLGKVTYFAPTYKDYDFSEDDYDSEKCRFPAMPDVLWM 810  
 RESULT 6  
 ULK1\_HUMAN  
 ID ULK1\_HUMAN STANDARD; PRT; 1050 AA.  
 AC 075385;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SERINE/THREONINE-PROTEIN KINASE ULK1 (EC 2.7.1.-) (UNC-51-LIKE KINASE  
 DE 1).  
 GN ULK1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98360094; PubMed=9693035;  
 RA Kuroyanagi H., Yan J., Seki N., Yamanouchi Y., Suzuki Y.I., Takano T.,  
 RA Muramatsu M.A., Shirasawa T.;  
 RT "Human ULK1, a novel serine/threonine kinase related to UNC-51 kinase  
 RT of Caenorhabditis elegans: cDNA cloning, expression, and chromosomal

```

RT assignment.
RL Genomics 51:76-85(1998).
CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED. DETECTED IN THE
CC FOLLOWING ADULT TISSUES: SKELETAL MUSCLE, HEART, PANCREAS, BRAIN,
CC PLACENTA, LIVER, KIDNEY, AND LUNG.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC APG1/UNC-51/DUK1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF045458; AAC32326.1; -
DR MIM: 603168; -
DR InterPro: IPR000719; -
DR InterPro: IPR001245; -
DR InterPro: IPR002290; -
DR Pfam: PF00069; PKinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Serine/threonine-protein kinase; ATP-binding.
KW TRANSFERASE; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 16 278 PROTEIN KINASE
FT NP_BIND 22 30 ATP (BY SIMILARITY).
FT BINDING 46 46 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
FT DOMAIN 297 310 POLY-SER.
SQ SEQUENCE 1050 AA; 112601 MW; 4ED9B94028E3C138 CRC64;

Query Match
Best Local Similarity 7.9%; Score 76.5; DB 1; Length 1050;
Matches 43; Conservative 13; Mismatches 47; Indels 39; Gaps 9;

DB 18 PLSGMRGIRGADFOCFQARAVGLSGTE-RAFLSLRLODLVSYVRA-----DGSVYP 70
DB 674 PLGPGLR--FGED-----PKGPGKRSFSTSRLLTDL--LTKAAGTQAPDPCSTE 718
OY 71 IYNLKDEVLSPSWDSLFGSGGOLQPGARI--FSDGRVLRHPAMPQKSVYH--GSDPSGR 129
DB 719 SIQEKPMETIAPS-----AGFGSLHRCAGAGTSS-----PSPVVFTVGSPPSGST 764
OY 130 LMEISYCEYRTETGATGQASS 151
DB 765 PPOG---PRTMFSAAGPTGSASS 784

RESULT 7
YHCC_ECOLI STANDARD: PRT; 309 AA.
AC P43476;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 34.6 KDA PROTEIN IN ARCB-GLTB INTERGENIC REGION.
GN YHCC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RC MEDLINE=97426517; PubMed=9278503;
RA Biltner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley J.M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

```

```

RT *The complete genome sequence of Escherichia coli K-12.*
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1-39 FROM N.A.
RA Ernsting B.R., Drenthinger J., Blumenthal R.M., Matthews R.G.;
RL Submitted (JUN-1993) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO M.JANNASCHII M10486.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18997; AAA58013.1; ALT_SEQ.
DR EMBL: AE000400; AAC76243.1; -
DR EMBL: L20253; -: NOT_ANNOTATED_CDS.
DR Ecogene: EG12809; Yhcc.
KW Hypothetical protein...
FT CONFLICT 24 25 KL-> NV (IN REF. 2).
SQ SEQUENCE 309 AA; 34606 MW; 61B3187B77CA1A9 CRC64;

Query Match
Best Local Similarity 7.9%; Score 76; DB 1; Length 309;
Matches 35; Conservative 23; Mismatches 42; Indels 34; Gaps 10;

DB 27 RGADFOCFQD---AR-----AVGASGTFRAFLSLRLODLVSYVRADGVS--P 70
DB 159 RGHDFPCYQRTTGLARNGKVCVSHLVGLPGSGA---ECLDTLRVYVETGVGDKLKP 215
OY 71 IYNLKDEVLSPSWDSLFGSGGOLQPGARI--FSDGRVLRHPAMPQKSVYH--GSDPSGR 128
DB 216 LHIYKGSIMAKWEA-----GRLN-GIELEDYTLTAGEMIRHT--PPEVIYHISASAR 266
OY 129 R---LMEISYCEY-T-W 138
DB 267 RPTLLAPLWCENRW 280

RESULT 8
SYJ1_HUMAN STANDARD: PRT; 1575 AA.
AC O43426; O43425;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHOSPHATE 5-
DE PHOSPHATASE 1).
GN SYNJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum.
RC MEDLINE=98088905; PubMed=9428629;
RA Haffner C., Takei K., Chen H., Ringstad N., Hudson A., Butler M.H.,
RA Salcini A.E., Di Fiore P.P., De Camilli P.;
RA "Synaptojanin 1: localization on coated endocytic intermediates in
RT nerve terminals and interaction of its 170 kDa isoform with Eps15";
RL FEBS Lett. 419:175-180(1997).
CC -1- FUNCTION: INOSITOL 5-PHOSPHATASE WHICH HAS A ROLE IN CLATHRIN-
CC MEDIATED ENDOCYTOSIS.
CC -1- CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O =
CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG ISOFORM/SYNAPTOJANIN-
CC 170 (SHOWN HERE) AND A SHORT ISOFORM/SYNAPTOJANIN-145; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: CONCENTRATED AT CLATHRIN-COATED ENDOCYTIC

```

```

CC INTERMEDIATES IN NERVE TERMINALS. THE LONG ISOFORM IS MORE
CC ENRICHED THAN THE SHORT ISOFORM IN DEVELOPING BRAIN AS WELL AS
CC NON-NEURONAL CELLS. THE SHORT ISOFORM IS VERY ABUNDANT IN NERVE
CC TERMINALS.
CC -1- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A
CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.
CC -1- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A
CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYLIN,
CC SH3P4 AND GRB2.
CC -1- SIMILARITY: IN THE CENTRAL SECTION, BELONGS TO THE INOSITOL-1,4,5-
CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SAC1 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC .....
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC OR SEND AN EMAIL TO license@sib-sib.ch).
CC .....
CC EMBL: AF009040; AAC51922.1; -.
CC EMBL: AF009039; AAC51921.1; -.
CC MIM: 604297; -.
CC InterPro: IPR000300; -.
CC DR PFAM: PF00783; IPPE; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC KM Hydrolyase; Alternative splicing; Repeat; Endocytosis; RNA-binding;
CC Multigene family.
CC FT DOMAIN 1 499 SAC1.
CC FT DOMAIN 500 899 CATALYTIC (POTENTIAL).
CC FT DOMAIN 902 971 RNA-BINDING (RRM).
CC FT DOMAIN 900 1575 PRO-RICH.
CC FT DOMAIN 1033 1036 POLY-SER.
CC FT DOMAIN 1108 1113 POLY-PRO.
CC FT DOMAIN 1126 1129 POLY-PRO.
CC FT DOMAIN 1487 1490 POLY-GLU.
CC FT DOMAIN 1540 1546 POLY-PRO.
CC FT DOMAIN 1396 1419 3 X 3 AA REPEATS OF N-P-F.
CC FT REPEAT 1396 1398 1.
CC FT REPEAT 1406 1408 2.
CC FT REPEAT 1417 1419 3.
CC FT VARSPLIC 1306 1311 VKRNGI -> QEQPSG (IN SHORT ISOFORM).
CC FT VARSPLIC 1312 1575 MISSING (IN SHORT ISOFORM).
CC SO SEQUENCE 1575 AA; 173345 MW; 506466CC043B9E7 CRC64;

Query Match 7.9%; Score 76; DB 1; Length 1575;
Best Local Similarity 22.6%; Pred. No. 27;
Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;

QY 8 PVLHVALNTPPLS--GGMRTGRCG-----DFOCFQOAR-AVGISGTFRAFLSSRLQ 55
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 648 PFIIDVAVDYVTKGMDGATGKGAVALRMLEHTTSLCFVCSHFAAGSQ-----VKERNE 702
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 56 DLXSVLRAD-----RGSPVTLNKDEVLSPMSDLFSGSQ--- 91
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 703 DFLFIARKLSPPMGRLPSHDYVFWCGDFNTRIDLPMNEVKEKLRLQNMDSLLINGDDQLIN 762
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 92 ----GQLQPG-----ARIFS--FDGNDVLRHPMPOKSVW 120
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 763 QKNAGVFRGFLGKVTFAPTYKYDLFSDDYDTSEKCRTPAMTRVLW 810
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 9
ID CATL_METBA STANDARD; PRT; 505 AA.
AC 093662;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATLASE (EC 1.11.1.6).

```

```

GN KAT.
OS Methanosarcina barkeri.
OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC Methanosarcina
OX NCBI_TaxID=2208;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=FUSARO / DSM 804;
RA Shima S., Netrusov A., Sordel M., Wicke M., Hartmann G.C.,
RA Thauer R.K.;
RT *Purification, characterization and primary structure of a
RT monofunctional catalase from Methanosarcina barkeri.*;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1- COFACTOR: HEME GROUP (BY SIMILARITY)
CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC .....
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC OR SEND AN EMAIL TO license@sib-sib.ch).
CC .....
CC EMBL: AJ005939; CAA06774.1; -.
CC HSP: P00432; 7CAT.
CC DR InterPro: IPR002226; -.
CC DR PFAM: PF00199; catalase; 1.
CC DR PRINTS: PR00067; CATALASE.
CC DR PROSITE: PS00437; CATALASE_1; 1.
CC DR PROSITE: PS00438; CATALASE_2; 1.
CC KM Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
CC FT ACT_SITE 58 58 BY SIMILARITY.
CC FT ACT_SITE 131 131 PROXIMAL HEME LIGAND (BY SIMILARITY).
CC FT BINDING 341 341 2A27C4BEC47BE854 CRC64;
CC SO SEQUENCE 505 AA; 57065 MW; 2A27C4BEC47BE854 CRC64;

Query Match 7.7%; Score 75; DB 1; Length 505;
Best Local Similarity 20.7%; Pred. No. 9;
Matches 30; Conservative 30; Mismatches 47; Indels 38; Gaps 7;

QY 35 QGARAVGLSGTFRAFLSSRLQDLYSLVRADRGSPVTLNKDEVLSPMSDLFSGSQGL 94
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 231 EEAEKIGSDPDHA-----TRDYEATK---KQDYPSWTLEMOIMPE----- 270
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 95 QPGARIPSPGDRVLRHPAMPKSVHMGSDPS--GRLMESYCEWRTETTGATGQASS 151
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 271 --QAEDYRFDIRDTK--VMP-----HGDEPTMIGKLVLRNPNTYFAVEQAALFSPAN 321
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 152 LLSC-----RLLEOKAASCHNSYI 170
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 322 LVPGIGISPDKMLQGRVFSYHDTI 346
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 10
ID HEMO_HUMAN STANDARD; PRT; 587 AA.
AC P22557;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 5-AMINOLEVULINIC ACID SYNTHASE, ERYTHROID-SPECIFIC, MITOCHONDRIAL
DE PRECURSOR (EC 2.3.1.37) (DELTA-AMINOLEVULINATE SYNTHASE) (DELTA-ALA
DE SYNTHETASE) (ALAS-E).
GN ALAS2 OR ALASE OR ASB.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```

Disease mutation. 749  
 KM TRANSIT 1  
 FT CHAIN 750 587  
 FT BINDING 391 391  
 FT VARIANT 388 388  
 FT VARIANT 411 411  
 FT VARIANT 476 476  
 FT VARIANT 476 476  
 SO SEQUENCE 587 AA; 64693 MM; DIA93AABE21CC794 CRC64;  
 Query Match 7.7%, Score 75; DB 1; Length 587;  
 Best Local Similarity 24.6%, Pred. No. 11;  
 Matches 49; Conservative 23; Mismatches 57; Indels 70; Gaps 10;  
 QY 49 FLSSLDPLYS-IVRRARGSVPIV-----NLKDEIVSPSWDSIFSSGQGLQ--- 95  
 DB 72 FWLSLDLQKSKTYVKA-----APEQEDVYKAFKTDLPSSLVSLRKPFSGPOEQIGS 127  
 QY 96 -----PCARIFSEDG--RDYLRI-----PAMP----- 115  
 DB 128 KYTHLIQNNMPGNYVSYDFFRDKIMEKQDHTYRYFKTVYRNADAPFQGHFFEASVA 187  
 QY 116 --QKSVHSGSPSGR---RLMESICGEVRFETGATGQASLLSGLR-----LEQKAAS 164  
 DB 188 SKDVSVMCSNDYLGSRHPQVLAQNETLRRHGAGAGCTRN--ISGTSKFFHELEQELAE 245  
 QY 165 CHNSTYVLCIENSFWTSFS 183  
 DB 246 LHKDSALLFSSCFVANDS 264  
 RESULT 11  
 SYTL\_BOVIN  
 ID SYTL\_BOVIN STANDARD; PRT; 1324 AA.  
 AC O18964;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPPIC INOSITOL-1,4,5-TRISPHOSPHATE 5-  
 DE PHOSPHATASE 1) (P150) (FRAGMENT).  
 GN SYNJ1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID:9913;  
 RN RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 321-339 AND 454-469.  
 RC TISSUE-Brain;  
 RX MEDLINE-97342621; PubMed-9199318;  
 RA Sakisaka T., Itoh T., Miura K., Takenawa T.;  
 RT "Phosphatidylinositol 4,5-bisphosphate phosphatase regulates the  
 RT rearrangement of actin filaments.";  
 RL M01. Cell. Biol. 17:3841-3849(1997).  
 CC -1- FUNCTION: HYDROLYSES PIP2 BOUND TO ACTIN REGULATORY PROTEINS  
 CC RESULTING IN THE REARRANGEMENT OF ACTIN FILAMENTS DOWNSTREAM OF  
 CC TYROSINE KINASE AND ASH/GRB2.  
 CC -1- CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O =  
 CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.  
 CC -1- SUBUNIT: BINDS TO AMPHIPHYLIN AND ASH/GRB2.  
 CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY CONCENTRATED IN THE  
 CC PERINUCLEAR AREAS.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGHEST LEVELS IN  
 CC BRAIN.  
 CC -1- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A  
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYLIN,  
 CC AND ASH/GRB2.  
 CC -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-

```

CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SAC1 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A NUMBER OF POTENTIAL
CC FRAMESHIFTS WERE CORRECTED STARTING IN POSITION 1213 SO AS TO
CC EXTEND THE SIMILARITY WITH THE ORTHOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: D85682; BAA21652.1; ALT_FRAME.
CC InterPro: IPR000300; -.
CC InterPro: IPR000504; -.
CC InterPro: IPR002013; -.
CC Pfam: PF00783; IPR01; -.
CC PROSITE: PS50102; RRM; 1.
CC Hydrolyase; Endocytosis; RNA-binding; Multigene family.
CC KW DOMAIN 1 474
CC FT DOMAIN 475 859
CC FT DOMAIN 902 971
CC FT DOMAIN 960 1212
CC FT DOMAIN 1033 1036
CC FT DOMAIN 1108 1113
CC FT DOMAIN 1126 1129
CC FT CONFLICT 1126 1129
CC FT CONFLICT 1324 1324
CC FT NON_TER 1324 1324
CC SQ SEQUENCE 1324 AA; 146582 MW; EDDC2DD9D6E3191C CRC64;

Query Match 7.78; Score 75; DB 1; Length 1324;
Best Local Similarity 22.68; Pred. No. 27;
Matches 38; Conservative 20; Mismatches 50; Indels 60; Gaps 8;

OY 8 PVLVALNTPLS*-GGMGRIGRA-----DFQFOQAR-AVSGTFRFAFLSSRLQ 55
DB 648 PFIDVAVDYKTMGATGKGNKGNALRMLEHTTSLCFVCSHPAAGSQ-----VKERN 702
OY 56 DLVSIVRRAD-----RGSVPVINKDEVLSFSDLSFGSQ--- 91
DB 703 DEFIAKRLSPMGRLFFSHDYVFCGDFNRYRIDLPNEVEKELRQGNMDSLGIAGDQLIN 762
OY 92 ----GQLQPG-----ARIS--FDGRDYLHRPAMPQKSW 120
DB 763 QKNAGOLFRRGLGKVTVPAPTYKYDLFSDDYDISEKCRTPAMTDRLVM 810

RESULT 12
PHBC_RHIME STANDARD: PRT; 611 AA.
AC P50176; 087321;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2.3.1.-) (POLY(3-
DE HYDROXYBUTYRATE) POLYMERASE) (PHB POLYMERASE) (PHB SYNTHASE)
DE (POLYHYDROXYALKANOIC ACID SYNTHASE).
GN PHBC.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
ON [1]
RP SEQUENCE OF 35-611 FROM N.A.
RC STRAIN=41;
RA MEDLINE=96036213; PubMed=7582015;
RA Tomboloni R., Povoledo S., Buson A., Squartini A., Nuti M.P.;
RA "Poly-beta-hydroxybutyrate (PHB) biosynthetic genes in Rhizobium

```

```

RT meliloti 41.";
RL Microbiology 141:2553-2559(1995).
RC [2]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=98405222; PubMed=9734305;
RA Willis L.B., Walker G.C.;
RA "The phbC (poly-beta-hydroxybutyrate synthase) gene of Rhizobium
RT (Sinorhizobium) meliloti and characterization of phbC mutants.";
RN Can. J. Microbiol. 44:554-564(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA Boistard P., Gouzy J., Kahn D., Thebaud P., Goffeau A.,
RA Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA Vandenberg M., Puehler A., Becker A., Weidner S.;
RL Submitted (MAR-2000) to the SWISS-PROT data bank.
CC -1- FUNCTION: POLYMERIZES D(-)-3-HYDROXYBUTYRYL-COA TO CREATE PHB
CC WHICH CONSISTS OF THOUSANDS OF HYDROXYBUTYRATE MOLECULES LINKED
CC WHEN CELLS GROW UNDER CONDITIONS OF NUTRIENT LIMITATION.
CC -1- PATHWAY: THIRD STEP IN POLY-BETA-HYDROXYBUTYRATE BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: U17227; AAA90984.1; -.
CC EMBL: AF031938; AAC61899.1; -.
CC InterPro: IPR000073; -.
CC Pfam: PF00561; abhydrolase; 1.
CC KW PHB biosynthesis; Transferase; Acyltransferase.
CC FT ACT SITE 349 349
CC FT CONFLICT 111 111
CC FT CONFLICT 122 122
CC FT CONFLICT 157 157
CC FT CONFLICT 477 477
CC FT CONFLICT 481 481
CC FT CONFLICT 481 481
CC SQ SEQUENCE 611 AA; 68039 MW; ABE8A9B17F87D7A9 CRC64;

Query Match 7.78; Score 74.5; DB 1; Length 611;
Best Local Similarity 27.48; Pred. No. 13;
Matches 40; Conservative 16; Mismatches 47; Indels 43; Gaps 9;

OY 49 FLSSRLQ-DLVSIVRRADGSV--PIVNL--KDEVFSPMSDLSFGSGQLPGARIFS 102
DB 470 YLENRLSKGEVLAGRVSLGVDYKIPLYNLATKEDHAPAKSYFLGSS-----S 518
OY 103 FDGR-----DVLHRPAMPQKSWHSGDSPSGRRLMESYCEW---RTETGA-- 145
DB 519 FGKAVTFVLSGSGHIAVVPAPRSKRYQWTGAPKGD-----IETWCKAKETAGSWM 572
OY 146 ----TGQASSLSGLRLLEOKAASCINS 168
DB 573 PHMOGWVERLDKRRVPARKKCGPLNS 598

RESULT 13
PFX_ECOLI STANDARD: PRT; 512 AA.
ID PFX_ECOLI
AC P29014;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)

```

DE EXOPOLYPHOSPHATASE (EC 3.6.1.11) (METAPHOSPHATASE).  
GN ppx.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
RC STRAIN-K12;  
RC MEDLINE=93107072; PubMed=8380170;  
RA Akiyama M., Crooke E., Kornberg A.;  
RT "An exopolyposphatase of Escherichia coli. The enzyme and its ppx  
gene in a polyphosphate operon.";  
RL J. Biol. Chem. 268:633-639(1993).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RC MEDLINE=97426617; PubMed=9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.M., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiiuchi T.,  
RA Ikemoto K., Inada T., Isono K., Itoh K., Kanai K., Kasai H.,  
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kikawa M., Makino K.,  
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,  
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,  
RA Tagami H., Takemoto K., Mada C., Yamamoto Y., Yano M.;  
RT Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RN SIMILARITY TO GPPA.  
RX MEDLINE=94025037; PubMed=8212131;  
RA Reizer J., Reizer A., Salier M.H. Jr., Bork B., Sander C.;  
RT "Exopolyposphate phosphatase and quanosine pentaphosphate  
phosphatase belong to the sugar kinase/actin/isp 70 superfamily.";  
RL Trends Biochem. Sci. 18:247-248(1993).  
CC -1- FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE  
IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA  
500 RESIDUES LONG, BUT CHAINS OF CIRCA 15 RESIDUES COMPLETE POORLY  
WITH POLYPHOSPHATE AS SUBSTRATE.  
CC -1- CATALYTIC ACTIVITY: (POLYPHOSPHATE)(N) + H(2)O =  
(POLYPHOSPHATE)(N-1) + ORTHOPHOSPHATE.  
CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (POTENTIAL).  
CC -1- SIMILARITY: TO E. COLI GUANOSINE-5'-TRIPHOSPHATE, 3'-DIPHOSPHATE  
PYROPHOSPHATASE (GPPA).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: L06129; AAA24415.1;  
DR EMBL: AE000336; AAC75555.1;  
DR EMBL: D90878; BAA16390.1;  
DR PIR: A45333; A45333.  
DR Ecogene: EG11403; ppx.  
KW Hydrolyase; Magnesium.  
FT NIT\_MET 0  
SQ SEQUENCE 512 AA; 58004 MW; 48611AF5D9FB9C3 CRC64;

Query Match 7.68; Score 74; DB 1; Length 512;  
Best Local Similarity 20.78; Pred. No. 11;

Matches 53; Conservative 40; Mismatches 77; Indels 86; Gaps 14;  
QY 4 QDFQVYL-----HLVALNTPLSGMRIGRADPQCFQAR----- 38  
DB 154 ENFEPILVESRRMGCVSFAQLTFP--GGV-----INKENFGARMAAOKLETLTWOPRI 206  
QY 39 -----AVGLSGTFR-----FLSSRLQDLX-STVRADAGSVIYNLKD- 77  
DB 207 QGWNVAMGASGITKAHVEYLMEMGEKDGITPERLEKLVKRLHRNPASLSLPLSEER 266  
QY 78 -----VLSPSWDSL-----FSGSOGLODPGARIFSFDRG-----DYLRRPAMPQKSW 120  
DB 267 KTVFVPGIALILCGVDALAIRLRISDGLARGV-LYMEGRFRQDYRSRASSLANQY 325  
QY 121 HCSDPGSRRLMES---YETWRTFTTG-ATGQASSLS-----GRLLEOKAASCHNS 168  
DB 326 HDSQARRVLDITMQMYEOWREQPKLAHQLLEALLRMAALHVEYGINNLSGLHRISA 385  
QY 169 YIVLCIENSFMFSRK 184  
DB 386 YI--LQNSDLPGFNO 398  
RESULT 14  
P47K.PSECL  
ID P47K.PSECL STANDARD: PRT: 419 AA.  
AC P31521;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 47 KDA PROTEIN (P47K).  
OS Pseudomonas chlororaphis (Pseudomonas fluorescens biotype D).  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=333;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-B23;  
RX MEDLINE=91193202; PubMed=2013568;  
RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,  
RA Beppu T.;  
RT "Cloning and characterization of genes responsible for metabolism of  
nitrile compounds from Pseudomonas chlororaphis B23.";  
RL J. Bacteriol. 173:2465-2472(1991).  
CC -1- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION  
OF THE NITRILE HYDRATASE GENES. MAY STABILIZE OR ACTIVATE THE  
NITRILE HYDRATASE PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: D90216; BAA1247.1;  
DR PIR: D42723; D42723.  
SQ SEQUENCE 419 AA; 46666 MW; FF5113800E27FF0C CRC64;

Query Match 7.58; Score 72.5; DB 1; Length 419;  
Best Local Similarity 22.18; Pred. No. 13;  
Matches 40; Conservative 24; Mismatches 60; Indels 57; Gaps 8;  
QY 26 IRCADPQCFQGARAVVGLSGTFRFLSSR-LQDLTSIVRADRGSVPIYNLKDDEVLSFSD 84  
DB 141 VDSQGFALLLESTDTVARADFEAHTSTRHLADL-LLEQVEYAVIIVNKRDLIDEGCYQ 198  
QY 85 SLFSGSOGLODPGARITS-----FDGRDYLIRHPAMPQK-----SVNHS 123  
DB 199 AVHAIIAG-LNPSARIMPMAHGYNVALSSLDLTHLFDPLSLAASGMRRKKEATDTDPASES 257

OY 124 DPGS-----RRMESYCEWRTTGTGATGQASSLLSGRLLEOKA-----ASC 165  
 DB 258 DTGCVTSWYRERAPRPPQRLLEFLQKPMH-----NGRLLRSKGYFWLASR 303  
 OY 166 H 166  
 DB 304 H 304

## RESULT 15

VNUC\_EBOG4  
 ID VNUC\_EBOG4 STANDARD: PRT: 739 AA.  
 AC Q9QCE9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NUCLEOPROTEIN (NUCLEOCAPSID PROTEIN).  
 GN NP.  
 OS Ebola virus (strain Gabon-94) (Ebo).  
 OC Viruses: ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
 OC Filovirus.  
 OX NCBI\_TaxID-128947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99036017; PubMed-9820131;  
 RA Prehaud C.J.C., Hellebrand E., Couderlier D., Volchkov V.E.,  
 RA Volchkova V.A., Feldmann B., Le Guenno B., Bouloy M.,  
 RT "Recombinant Ebola virus nucleoprotein and glycoprotein (Gabon 94  
 RT strain) provide new tools for the detection of human infections.";  
 RT J. Gen. Virol. 79:2565-2572(1998).  
 CC -1- FUNCTION: RESPONSIBLE FOR ENCAPSIDATION OF GENOMIC RNA.  
 CC -1- DOMAIN: THIS PROTEIN CAN BE DIVIDED INTO A HYDROPHOBIC N-TERMINAL  
 CC HALF, AND A HYDROPHILIC AND HIGHLY ACIDIC C-TERMINAL HALF.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: Y09358; CAA70541.1; -  
 KM Nucleoprotein.  
 SO SEQUENCE 739 AA: 83240 MW: 33A96720FBC638E CRC64;

## Query Match

Best Local Similarity 7.5%; Score 72.5; DB 1; Length 739;

Matches 43; Conservative 16; Mismatches 67; Indels 33; Gaps 8;

OY 22 GMRGIRGAD-----PQCFQARAVGLSGTFRALSSRLD-----LYSIVRRADRGSV 69  
 DB 218 GMHVVAGHDANDAVISNSVAQAPSGLL-IVKVLDIHLQKTQGVRIHPLARTA----- 271  
 OY 70 PIVNLKDEV--LSPSWDSLFSGSOGLOPGARIFSPGDRVLRHPAMPKRS-----VW 120  
 DB 272 ---KVKNEVNSLKAALSL--AKHGEYAPFARLLNLSGVNLEHGLFPQLSAIALGVATA 326  
 OY 121 HGSDDSGRRLMESYCEWRTTGTGATGQASSLLSGRLLE 159  
 DB 327 HGSLLAGVNVGEQY-QQLREAYTEAKQLOQYAESRELD 364

Search completed: August 28, 2001, 12:52:40  
 Job time: 109 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:49:31 ; Search time 12.29 Seconds  
(without alignments)  
308.269 Million cell updates/sec

Title: US-09-589-777a-2  
Perfect score: 968  
Sequence: 1 HTHODPQVHLVALNTPLS.....CHNSYVLCIENSFMTFSK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	965	99.7	195 1 US-08-159-784-2	Sequence 2, Appli
2	916	94.6	185 3 US-08-985-526-36	Sequence 36, Appl
3	840	86.8	183 4 US-09-206-059-2	Sequence 2, Appli
4	550	56.8	191 1 US-08-159-784-3	Sequence 3, Appli
5	150	15.5	35 3 US-09-046-985-2	Sequence 2, Appli
6	150	15.5	35 4 US-09-474-743-2	Sequence 2, Appli
7	108	11.2	20 2 US-08-740-168A-1	Sequence 1, Appli
8	108	11.2	20 4 US-09-349-429-1	Sequence 1, Appli
9	101	10.4	16 4 US-09-385-442-32	Sequence 32, Appli
10	101	10.4	22 3 US-09-046-985-7	Sequence 7, Appli
11	101	10.4	22 4 US-09-474-743-7	Sequence 7, Appli
12	75	7.7	190 3 US-09-046-985-15	Sequence 15, Appli
13	75	7.7	190 4 US-09-474-743-15	Sequence 15, Appli
14	75	7.7	587 2 US-08-871-266B-18	Sequence 18, Appli
15	75	7.7	587 2 US-09-018-864A-18	Sequence 18, Appli
16	75	7.7	587 2 US-08-871-267B-24	Sequence 24, Appli
17	74.5	7.7	577 2 US-08-756-317-13	Sequence 13, Appli
18	71.5	7.4	1214 2 US-08-231-193A-54	Sequence 54, Appli
19	71.5	7.4	1214 2 US-08-486-273A-54	Sequence 54, Appli
20	71.5	7.4	1214 3 US-08-480-474-54	Sequence 54, Appli
21	71.5	7.4	1214 3 US-08-940-086A-54	Sequence 54, Appli
22	71.5	7.4	1219 2 US-08-231-193A-50	Sequence 50, Appli
23	71.5	7.4	1219 2 US-08-486-273A-50	Sequence 50, Appli
24	71.5	7.4	1219 2 US-08-480-474-50	Sequence 50, Appli
25	71.5	7.4	1219 3 US-08-940-086A-50	Sequence 50, Appli
26	71.5	7.4	1231 2 US-08-231-193A-48	Sequence 48, Appli
27	71.5	7.4	1231 2 US-08-486-273A-48	Sequence 48, Appli

28	71.5	7.4	1231 3 US-08-480-474-48	Sequence 48, Appli
29	71.5	7.4	1231 3 US-08-940-086A-48	Sequence 48, Appli
30	71.5	7.4	1236 2 US-08-231-193A-6	Sequence 6, Appli
31	71.5	7.4	1236 2 US-08-486-273A-6	Sequence 6, Appli
32	71.5	7.4	1236 3 US-08-480-474-6	Sequence 6, Appli
33	71.5	7.4	1236 3 US-08-940-086A-6	Sequence 6, Appli
34	71.5	7.4	1239 2 US-08-231-193A-52	Sequence 52, Appli
35	71.5	7.4	1239 2 US-08-486-273A-52	Sequence 52, Appli
36	71.5	7.4	1239 3 US-08-480-474-52	Sequence 52, Appli
37	71.5	7.4	1239 3 US-08-940-086A-52	Sequence 52, Appli
38	71.5	7.4	1244 2 US-08-231-193A-46	Sequence 46, Appli
39	71.5	7.4	1244 2 US-08-486-273A-46	Sequence 46, Appli
40	71.5	7.4	1244 3 US-08-480-474-46	Sequence 46, Appli
41	71.5	7.4	1244 3 US-08-940-086A-46	Sequence 46, Appli
42	71	7.3	1289 1 US-07-876-280-4	Sequence 4, Appli
43	71	7.3	1289 1 US-07-675-170-4	Sequence 4, Appli
44	71	7.3	1289 1 US-08-053-170-4	Sequence 4, Appli
45	71	7.3	1289 1 US-08-158-232-4	Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-159-784-2  
: Sequence 2, Application US/08159784  
: Patent No. 5643783  
: GENERAL INFORMATION:  
: APPLICANT: Bjorn R. Olsen  
: TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Fish & Richardson  
: STREET: 225 Franklin Street  
: CITY: Boston  
: STATE: Massachusetts  
: COUNTRY: U.S.A.  
: ZIP: 02110-2804  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
: COMPUTER: IBM PS/2 Model 502 or 555X  
: OPERATING SYSTEM: MS-DOS (Version 5.0)  
: SOFTWARE: Wordperfect (Version 5.1)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/159,784  
: FILING DATE: December 1, 1993  
: CLASSIFICATION: 530  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: John F. Freeman  
: REGISTRATION NUMBER: 29,066  
: REFERENCE/DOCKET NUMBER: 00246/170001  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 542-5070  
: TELEFAX: (617) 542-8906  
: TELETYPE: 200154  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 195  
: TYPE: amino acid  
: STRANDEDNESS: N/A  
: TOPOLOGY: N/A  
: US-08-159-784-2

Query Match 99.7%: Score 965; DB 1; Length 195;  
Best Local Similarity 99.5%: Pred. No. 5.4e-112;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 HTHODPQVHLVALNTPLSGGMGIRGADFCQOARAVGISTGFRAFLSSRLDLYSI 60  
|||||

Db 12 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 71  
 Oy 61 VRRADGSPVILVKDEVLSPSWDSLFSGSGQLOLPGARIFSPDGRDYLRRHPAMPQKSV 120  
 Db 72 VRRADGSPVILVKDEVLSPSWDSLFSGSGQLOLPGARIFSPDGRDYLRRHPAMPQKSV 131  
 Oy 121 HGSDFSGRRLMESYCEWTWRTETTGATGQASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180  
 Db 132 HGSDFSGRRLMESYCEWTWRTETTGATGQASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 191  
 Oy 181 SFSK 184  
 Db 192 SFSK 195

RESULT 2  
 US-08-985-526-36  
 Sequence 36, Application US/08985526  
 Patent No. 6080728

GENERAL INFORMATION:  
 APPLICANT: Mixson, James A  
 TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA  
 TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
 TITLE OF INVENTION: THERAPY  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Connolly, Bove, Lodge, & Hutz  
 STREET: 1220 Market Street, P.O. Box 2207  
 CITY: Wilmington  
 STATE: Delaware  
 COUNTRY: U.S.A.  
 ZIP: 19899  
 COMPUTER TYPE: FLOPPY disk  
 MEDIUM TYPE: PC-DOS/MS-DOS  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/985,526  
 FILING DATE:  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/608,845  
 FILING DATE: 16-JUL-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: McGorrow Jr., Robert G  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (302) 658-9141  
 TELEFAX: (302) 658-5613  
 INFORMATION FOR SEQ ID NO: 36:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 185 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-985-526-36

Query Match 94.6%; Score 916; DB 3; Length 185;  
 Best Local Similarity 95.1%; Pred. No. 6, 1e-106;  
 Matches 176; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

Oy 1 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
 Db 2 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
 Oy 61 VRRADGSPVILVKDEVLSPSWDSLFSGSGQLOLPGARIFSPDGRDYLRRHPAMPQKSV 119  
 Db 61 VRRADGSPVILVKDEVLSPSWDSLFSGSGQLOLPGARIFSPDGRDYLRRHPAMPQKSV 120  
 Oy 120 HGSDFSGRRLMESYCEWTWRTETTGATGQASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 179  
 Db 121 HGSDFSGRRLMESYCEWTWRTETTGATGQASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180

Oy 180 TSFSK 184  
 Db 181 TSFSR 185

RESULT 3  
 US-09-206-059-2  
 Sequence 2, Application US/09206059  
 Patent No. 6201104  
 GENERAL INFORMATION:  
 APPLICANT: Macdonald, Nicholas  
 APPLICANT: Sim, Kim Lee  
 TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and  
 TITLE OF INVENTION: Proteins and Methods of Use  
 FILE REFERENCE: 05213-0370  
 CURRENT APPLICATION NUMBER: US/09/206,059  
 CURRENT FILING DATE: 1998-12-04  
 NUMBER OF SEQ ID NOS: 80  
 SOFTWARE: Patent Ver. 2.0  
 SEQ ID NO 2  
 LENGTH: 183  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-206-059-2

Query Match 85.8%; Score 840; DB 4; Length 183;  
 Best Local Similarity 85.6%; Pred. No. 1, 7e-96;  
 Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Oy 1 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
 Db 1 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
 Oy 61 VRRADGSPVILVKDEVLSPSWDSLFSGSGQLOLPGARIFSPDGRDYLRRHPAMPQKSV 120  
 Db 61 VRRADGSPVILVKDEVLSPSWDSLFSGSGQLOLPGARIFSPDGRDYLRRHPAMPQKSV 120  
 Oy 121 HGSDFSGRRLMESYCEWTWRTETTGATGQASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180  
 Db 121 HGSDFSGRRLMESYCEWTWRTETTGATGQASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180  
 Oy 181 S 181  
 Db 181 A 181

RESULT 4  
 US-08-159-784-3  
 Sequence 3, Application US/08159784  
 Patent No. 5643783  
 GENERAL INFORMATION:  
 APPLICANT: Bjorn R. Olsen  
 TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Fish & Richardson  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
 COMPUTER: IBM PS/2 Model 502 or 555X  
 OPERATING SYSTEM: MS-DOS (Version 5.0)  
 SOFTWARE: WordPerfect (Version 5.1)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/159,784  
 FILING DATE: December 1, 1993  
 CLASSIFICATION: 530  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER:

STRANDEDNESS:

; GENERAL INFORMATION:

RESULT 7  
US-08-740-168A-1  
; Sequence 1, Application US/08740168A  
; Patent No. 5854205  
; GENERAL INFORMATION:  
; APPLICANT: O'Reilly, Michael

APPLICANT: Folkman, M. Judeh  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askey, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168A  
FILING DATE: 22-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-08-740-168A-1

Query Match 11.2% Score 108 DB 2: Length 20:  
Best Local Similarity 100.0% Pred. No. 5.5e-07:  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 HTHODFQPVHLVALNTPLS 20  
DB 1 HTHODFQPVHLVALNTPLS 20

RESULT 8  
US-09-349-429-1  
Sequence 1, Application US/09349429  
Patent No. 6174861  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, Michael  
APPLICANT: Folkman, M. Judeh  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askey, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/349,429  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168  
FILING DATE: 22-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-09-349-429-1

Query Match 11.2% Score 108 DB 4: Length 20:  
Best Local Similarity 100.0% Pred. No. 5.5e-07:  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 HTHODFQPVHLVALNTPLS 20  
DB 1 HTHODFQPVHLVALNTPLS 20

RESULT 9  
US-09-385-442-32  
Sequence 32, Application US/09385442  
Patent No. 6200954  
GENERAL INFORMATION:  
APPLICANT: Ge, Ruowen  
APPLICANT: Kini, R. Manjunatha  
TITLE OF INVENTION: Small Peptides Having Potent Anti-Angiogenic Activity  
FILE REFERENCE: 1781-170P  
CURRENT APPLICATION NUMBER: US/09/385,442  
CURRENT FILING DATE: 1999-08-30  
EARLIER APPLICATION NUMBER: 60/099,313  
EARLIER FILING DATE: 1999-09-04  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 32  
LENGTH: 16  
TYPE: PRT  
ORGANISM: mammalian  
FEATURE:  
OTHER INFORMATION: Endo-4  
US-09-385-442-32

Query Match 10.4% Score 101 DB 4: Length 16:  
Best Local Similarity 100.0% Pred. No. 2.9e-06:  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 111 HPAPPOKSVWHGSDPS 126  
DB 1 HPAPPOKSVWHGSDPS 16

RESULT 10  
US-09-046-985-7  
Sequence 7, Application US/09046985  
Patent No. 6121236  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,985  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-9540  
TELEFAX: (781) 861-6240  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= modified aa  
US-09-046-985-7

Query Match 10.4%; Score 101; DB 3; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 TFRFLSSRLDLYSIVRRAD 65  
|||||  
DB 1 TFRFLSSRLDLYSIVRRAD 21

RESULT 11  
US-09-474-743-7  
Sequence 7, Application US/09474743  
Patent No. 6235716  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/474,743  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/046,985  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240  
TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= modified aa  
US-09-474-743-7

Query Match 10.4%; Score 101; DB 4; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 TFRFLSSRLDLYSIVRRAD 65  
|||||  
DB 1 TFRFLSSRLDLYSIVRRAD 21

RESULT 12  
US-09-046-985-15  
Sequence 15, Application US/09046985  
Patent No. 6121236  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,985  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240

TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-046-985-15

Query Match 7.7%; Score 75; DB 3; Length 190;  
Best Local Similarity 44.7%; Pred. No. 0.21;  
Matches 21; Conservative 2; Mismatches 12; Indels 12; Gaps 1;

OY 21 GGMGIRGADPFCQOARAVGLSGTFRALFSSRLQDLYSTRADRG 67  
DB 34 GSGGGGSKDFT-----AYRWRLSHRPKDLSTRADRG 68

RESULT 13  
US-09-474-743-15  
Sequence 15, Application US/09474743  
Patent No. 6235716  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/474,743  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/046,985  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240  
TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-474-743-15

Query Match 7.7%; Score 75; DB 4; Length 190;  
Best Local Similarity 44.7%; Pred. No. 0.21;  
Matches 21; Conservative 2; Mismatches 12; Indels 12; Gaps 1;

OY 21 GGMGIRGADPFCQOARAVGLSGTFRALFSSRLQDLYSTRADRG 67  
DB 34 GSGGGGSKDFT-----AYRWRLSHRPKDLSTRADRG 68

RESULT 14  
US-08-871-266B-18  
Sequence 18, Application US/08871266B  
Patent No. 5871991  
GENERAL INFORMATION:  
APPLICANT: Elirod, Susan L.  
TITLE OF INVENTION: Aspergillus Oryzae 5-Aminolevulinic  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 58719910 No. 5871991disk Of No. 5871991th America, Inc.  
STREET: 405 Lexington Avenue - 64th Fl.  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871.266B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambitis, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-878-9652  
TELEFAX: 212-878-9655  
TELEX:

INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 587 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5871991e  
US-08-871-266B-18

Query Match 7.7%; Score 75; DB 2; Length 587;  
Best Local Similarity 24.6%; Pred. No. 1.1;  
Matches 49; Conservative 23; Mismatches 57; Indels 70; Gaps 10;

OY 49 FLSSRLQDLYS-IVRRADRGVPIV-----NKDEVLSPSWDSLFGSGGQDQ--- 95  
DB 72 FMSLQDQSKSIYOKA-----APEVQEDVYKAFKTDLPSSLVSLSLRPFSGPOEEOISG 127  
OY 96 -----PGARISFDG-RDYLRH-----PAMP----- 115  
DB 128 KYTHLQNNMNGVVSVDQFPRDKIMEKKQDHTYRVKTVNMWADAPPAQHFEEASVA 187  
OY 116 --QKSVWHSQDPGSR-----RLMESYCEWTWTETTGATGQASLSLGR-----LEQKAAS 164  
DB 188 SKDVSYWCSDNYLGMSRHPVLDATQETLDRHAGAGGTNN--ISGTSKHVLEEDQLAE 245  
OY 165 CHNSYIVLCIENSEFMTSFS 183  
DB 246 LHKQDSALLFSSCFVANDS 264

RESULT 15  
US-09-018-864A-18  
Sequence 18, Application US/09018864A  
Patent No. 5958747  
GENERAL INFORMATION:  
APPLICANT: Elirod, Susan L.  
APPLICANT: Cherry, Joel R.





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 28, 2001, 12:50:31 ; Search time 22.57 Seconds  
(without alignments)  
1078.605 Million cell updates/sec

Title: US-09-589-777A-2  
Perfect score: 968  
Sequence: 1 HTHODFQPVHLVLAALNTPLS.....CHNSYIVLCIENSFTSFSK 184  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*\*\*  
2: SP-bacteria:\*\*\*  
3: SP-fungi:\*\*\*  
4: SP-human:\*\*\*  
5: SP-invertebrate:\*\*\*  
6: SP-mammal:\*\*\*  
7: SP-mhc:\*\*\*  
8: SP-organelle:\*\*\*  
9: SP-phage:\*\*\*  
10: SP-plant:\*\*\*  
11: SP-prodent:\*\*\*  
12: SP-unclassified:\*\*\*  
13: SP-vertebrate:\*\*\*  
14: SP-virus:\*\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	968	100.0	184	11	Q9JK63 mus musculu
2	968	100.0	1140	11	Q61434 mus musculu
3	968	100.0	1774	11	Q62001 mus musculu
4	927	95.8	1226	11	Q9QZD2 rattus norv
5	858	88.6	1774	11	Q9WJMS rattus norv
6	754	77.9	386	13	Q93419 gallus gall
7	574	59.3	1367	11	Q35206 mus musculu
8	573	59.2	1367	11	Q9EQD9 mus musculu
9	564	58.3	1388	4	Q9Y4M4 mus musculu
10	425.5	44.0	581	5	Q9VSD9 drosophila
11	376	38.8	650	5	Q17866 caenorhabdi
12	376	38.8	778	5	Q9U9K6 caenorhabdi
13	376	38.8	1117	5	Q9U9K7 caenorhabdi
14	85	8.8	995	4	Q9Y3M8 homo sapien
15	83.5	8.6	651	5	Q9VFA9 drosophila
16	82	8.5	483	5	Q9U5L0 subterites d
17	81	8.4	539	2	Q914Y1 streptomyc
18	79	8.2	427	11	Q62700 rattus norv
19	79	8.2	1112	14	Q9YUR6 turkey aden

20	78.5	8.1	416	2	Q9L609 salmonella
21	78.5	8.1	9477	2	Q914X3 streptomyc
22	78	8.1	396	5	Q9YDN3 drosophila
23	78	8.1	1055	10	Q39725 euglena gra
24	77.5	8.0	474	2	Q50185 mycobacteri
25	77.5	8.0	6420	2	P95814 streptomyc
26	77	8.0	208	2	Q9X9V1 streptomyc
27	76.5	7.9	235	7	Q9NNM8 oncorhynch
28	76.5	7.9	784	4	Q9UQ28 homo sapien
29	76.5	7.9	1054	2	Q921X4 pseudocater
30	76	7.9	1193	4	Q94984 homo sapien
31	75.5	7.8	1136	10	Q9LW5 oryza sativ
32	75.5	7.8	4924	3	Q9E7T1 scilzosacch
33	75	7.7	587	4	Q13735 homo sapien
34	74.5	7.7	167	7	Q46771 bos taurus
35	74.5	7.7	436	5	Q9YAK0 drosophila
36	74.5	7.7	3670	2	Q924X5 streptomyc
37	74.5	7.7	7576	2	Q92G44 streptomyc
38	74	7.6	244	2	Q53235 rhodospacer
39	74	7.6	497	2	P76981 escherichia
40	74	7.6	517	14	Q83378 rat leukem
41	74	7.6	571	2	Q9L0Q4 streptomyc
42	74	7.6	1715	6	Q9GLM4 bos taurus
43	74	7.6	4180	2	Q915N6 pseudomonas
44	74	7.6	6396	2	Q9KID7 streptomyc
45	73.5	7.6	305	10	Q9M3R8 arabidopsis

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	184 AA.
Q9JK63	1			
AC	Q9JK63:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	ENDOSTATIN (FRAGMENT).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-CHINESE KUNMING;			
RA	Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;			
RT	*Anticancer treatment of targeted fusion protein delivery to tumor			
RT	neovascularization.*;			
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF257775; AAF69009.1; ..			
FT	NON-TER			
FT	NON-TER			
FT	NON-TER			
SQ	SEQUENCE 184 AA; 20376 MW; AC06998D103412A CRC64;			
Query Match	100.0%; Score 968; DB 11; Length 184;			
Best Local Similarity	100.0%; Pred. No. 71e-90; Indels 0; Gaps 0;			
Matches 184; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	HTHODFQPVHLVLAALNTPLSGMRCIRGADFCFOARAVGHSFTRAFISRLDLSI	60	Q916Q9 salmonella
DB	1	HTHODFQPVHLVLAALNTPLSGMRCIRGADFCFOARAVGHSFTRAFISRLDLSI	60	Q914X3 streptomyc
QY	61	VRADRGSPVIVNKLDEVLSPSMDSLFSGSGQLPGARIFSFDGRDYLRRHPAMPQKSVW	120	Q9YDN3 drosophila
DB	61	VRADRGSPVIVNKLDEVLSPSMDSLFSGSGQLPGARIFSFDGRDYLRRHPAMPQKSVW	120	Q39725 euglena gra
QY	121	HSDPSGRRLMSYCEMTRETTGATGQASLLSRLLDQRAASHNSYIVLCIENSFT	180	Q50185 mycobacteri
DB	121	HSDPSGRRLMSYCEMTRETTGATGQASLLSRLLDQRAASHNSYIVLCIENSFT	180	P95814 streptomyc
QY	181	SFSK 184		Q9X9V1 streptomyc
DB	181	SFSK 184		Q9NNM8 oncorhynch
QY	181	SFSK 184		Q9UQ28 homo sapien
DB	181	SFSK 184		Q921X4 pseudocater
QY	181	SFSK 184		Q94984 homo sapien
DB	181	SFSK 184		Q9LW5 oryza sativ
QY	181	SFSK 184		Q9E7T1 scilzosacch
DB	181	SFSK 184		Q13735 homo sapien
QY	181	SFSK 184		Q46771 bos taurus
DB	181	SFSK 184		Q9YAK0 drosophila
QY	181	SFSK 184		Q924X5 streptomyc
DB	181	SFSK 184		Q92G44 streptomyc
QY	181	SFSK 184		Q53235 rhodospacer
DB	181	SFSK 184		P76981 escherichia
QY	181	SFSK 184		Q83378 rat leukem
DB	181	SFSK 184		Q9L0Q4 streptomyc
QY	181	SFSK 184		Q9GLM4 bos taurus
DB	181	SFSK 184		Q915N6 pseudomonas
QY	181	SFSK 184		Q9KID7 streptomyc
DB	181	SFSK 184		Q9M3R8 arabidopsis

Db 181 SFSK 184

## RESULT 2

Q61434 PRELIMINARY; PRT: 1140 AA.

AC 061434  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last annotation update)  
 DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)  
 DE COLLAGEN (FRAGMENT).  
 GN COL15A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94059075; PubMed-8240330;  
 RA Abe N., Muragaki Y., Yoshida H., Inoue H., Niomura Y.;  
 RT Identification of a novel collagen chain represented by extensive  
 RT interruptions in the triple-helical region.  
 RL Cell. Mol. Biol. Res. 196:576-582(1993).  
 DR EMBL: D17546; BAA04483.1; -  
 DR HSSP: P39061; 1KOE.  
 DR MGD: MGI:88449; Col15a1.  
 DR InterPro: IPR000087; -  
 DR Pfam: PF01391; Collagen; 6.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDFE CRC64;

Query Match 100.0%; Score 968; DB 11; Length 1140;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-89;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVLTALNTPLSGCMRGIRGADFCQCARAVGSGTFRAPLSRLDLYSI 60  
 Db 957 HTHQDFQPVHLVLTALNTPLSGCMRGIRGADFCQCARAVGSGTFRAPLSRLDLYSI 1016  
 QY 61 VRRADGSPVIVLKDENVLSPSWDSLFGSGGOLQPCARIFSDGNDVLRHPAMPKQSVW 120  
 Db 1017 VRRADGSPVIVLKDENVLSPSWDSLFGSGGOLQPCARIFSDGNDVLRHPAMPKQSVW 1076  
 QY 121 HGSDDPGSRRLMESYCEWTWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIENSFMT 180  
 Db 1077 HGSDDPGSRRLMESYCEWTWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIENSFMT 1136  
 QY 181 SFSK 184  
 Db 1137 SFSK 1140

## RESULT 3

Q62001 PRELIMINARY; PRT: 1774 AA.

AC 062001  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last annotation update)  
 DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)  
 DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN  
 DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).  
 GN COL18A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94243707; PubMed-8188673;  
 RA Rehn M., Hintikka E., Pihlajaniemi T.;

RT Primary structure of the alpha 1 chain of mouse type XVIII collagen,  
 RT partial structure of the corresponding gene, and comparison of the  
 RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen  
 RT chain.  
 RL J. Biol. Chem. 269:13929-13935(1994).

RP SEQUENCE OF 1-562 FROM N.A.  
 RX MEDLINE-94240112; PubMed-8183894;  
 RA Rehn M., Pihlajaniemi T.;  
 RT "Alpha 1(XVIII)", a collagen chain with frequent interruptions in the  
 RT collagenous sequence, a distinct tissue distribution, and homology  
 RT with type XV collagen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).

RN [3]  
 RP SEQUENCE OF 1-562 FROM N.A.  
 RX MEDLINE-95181468; PubMed-7876242;  
 RA Rehn M., Pihlajaniemi T.;  
 RT Identification of three N-terminal ends of type XVIII collagen chains  
 RT and tissue-specific differences in the expression of the corresponding  
 RT transcripts. The longest form contains a novel motif homologous to rat  
 RT and Drosophila frizzled proteins."  
 RL J. Biol. Chem. 270:4705-4711(1995).

DR EMBL: U03715; AAC52903.1; -  
 DR EMBL: U03716; AAC52903.1; JOINED.  
 DR EMBL: U03718; AAC52903.1; JOINED.  
 DR EMBL: U34607; AAC52903.1; JOINED.  
 DR EMBL: U34608; AAC52903.1; JOINED.  
 DR EMBL: U34609; AAC52903.1; JOINED.  
 DR EMBL: U34610; AAC52903.1; JOINED.  
 DR EMBL: U34611; AAC52903.1; JOINED.  
 DR EMBL: U34612; AAC52903.1; JOINED.  
 DR EMBL: U34613; AAC52903.1; JOINED.  
 DR EMBL: U11637; AAC52179.1; -  
 DR HSSP: P39061; 1KOE.

DR MGD: MGI:88451; Col18a1.  
 DR InterPro: IPR000024; -  
 DR InterPro: IPR000087; -  
 DR InterPro: IPR001791; -  
 DR InterPro: IPR003129; -  
 DR Pfam: PF01392; Fz; 1.  
 DR Pfam: PF02210; TSPN; 1.  
 DR SMART: SM00282; LamG; 1.  
 KW Signal.  
 SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 968; DB 11; Length 1774;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-88;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVLTALNTPLSGCMRGIRGADFCQCARAVGSGTFRAPLSRLDLYSI 60  
 Db 1591 HTHQDFQPVHLVLTALNTPLSGCMRGIRGADFCQCARAVGSGTFRAPLSRLDLYSI 1650  
 QY 61 VRRADGSPVIVLKDENVLSPSWDSLFGSGGOLQPCARIFSDGNDVLRHPAMPKQSVW 120  
 Db 1651 VRRADGSPVIVLKDENVLSPSWDSLFGSGGOLQPCARIFSDGNDVLRHPAMPKQSVW 1710  
 QY 121 HGSDDPGSRRLMESYCEWTWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIENSFMT 180  
 Db 1711 HGSDDPGSRRLMESYCEWTWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIENSFMT 1770  
 QY 181 SFSK 184  
 Db 1771 SFSK 1774

## RESULT 4

Q90ZD2 PRELIMINARY; PRT: 226 AA.

AC 090ZD2  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
 DE COLLAGEN XVIII (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RA Chen L., Perletti G., Folkman J.;  
 RT "Antitumor activity of rat endostatin."  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF189709; AAF00975.1; -  
 DR HSSP; P39061; IKOE.  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 226 AA; 25350 MW; 38883C0486C0E949 CRC64;

Query Match 95.8%; Score 927; DB 11; Length 226;  
 Best Local Similarity 95.7%; Pred. No. 1.2e-85;  
 Matches 176; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 HTHODFQPVHLVALNTPLSGMGRIGRAGDFQCFQOARAVGLSGTFRAFLSSRLDLYSI 60  
 DB 43 HTHODFQPVHLVALNTPLSGMGRIGRAGDFQCFQOARAVGLSGTFRAFLSSRLDLYSI 102  
 OY 61 VRRADGSGVPIVNLKDEVLSFSSGQGLQPGARIFSPDGRVLRHPAMPKSVW 120  
 DB 103 VRRADGSGVPIVNLKDEVLSFSSGQGLQPGARIFSPDGRVLRHPAMPKSVW 162  
 OY 121 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 180  
 DB 163 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 222  
 OY 181 SFSK 184  
 DB 223 SFSK 226  
 RESULT 5  
 O9MDWS PRELIMINARY; PRT; 171 AA.  
 AC 09MDWS;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
 DE COLLAGEN TYPE XVIII, ALPHA (I) CHAIN (FRAGMENT).  
 GN COL18A1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Jia J.D., Bauer M., Eberspacher U., Donner P., Schuppan D.;  
 RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin."  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;  
 RT "Temporospatial expression of collagen XVIII/endostatin in acute and chronic liver injuries."  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ236873; CAB44263.1; -  
 DR HSSP; P39061; IKOE.  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match 88.6%; Score 856; DB 11; Length 171;  
 Best Local Similarity 95.9%; Pred. No. 7.9e-79;  
 Matches 164; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 9 VLHLVALNTPLSGMGRIGRAGDFQCFQOARAVGLSGTFRAFLSSRLDLYSIYRRADRS 68  
 DB 1 VLHLVALNTPLSGMGRIGRAGDFQCFQOARAVGLSGTFRAFLSSRLDLYSIYRRADRS 60  
 OY 69 VPIVNLKDEVLSFSSGQGLQPGARIFSPDGRVLRHPAMPKSVWHSDDPSGR 128  
 DB 61 VPIVNLKDEVLSFSSGQGLQPGARIFSPDGRVLRHPAMPKSVWHSDDPSGR 120  
 OY 129 RLMESECEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 179  
 DB 121 RLMESECEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 171

## RESULT 6

O93419 PRELIMINARY; PRT; 386 AA.  
 AC 093419;

DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)  
 DE COLLAGEN XVIII (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Haller W., Dong S., Schurer B., Cole G.;  
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan."  
 RL J. Biol. Chem. 0:0-0(1998).  
 DR EMBL; AF083440; AAC33294.1; -  
 DR HSSP; P39061; IKOE.  
 DR InterPro; IPR000087; -  
 FT NON\_TER  
 SQ SEQUENCE 386 AA; 41775 MW; 34D40FA9E8A3B0E CRC64;

Query Match 77.9%; Score 754; DB 13; Length 386;  
 Best Local Similarity 75.5%; Pred. No. 6.6e-68;  
 Matches 139; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

OY 1 HTHODFQPVHLVALNTPLSGMGRIGRAGDFQCFQOARAVGLSGTFRAFLSSRLDLYSI 60  
 DB 203 HTHODFQPVHLVALNTPLSGMGRIGRAGDFQCFQOARAVGLSGTFRAFLSSRLDLYSI 262  
 OY 61 VRRADGSGVPIVNLKDEVLSFSSGQGLQPGARIFSPDGRVLRHPAMPKSVW 120  
 DB 263 VRRADGSGVPIVNLKDEVLSFSSGQGLQPGARIFSPDGRVLRHPAMPKSVW 322  
 OY 121 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 180  
 DB 323 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 382  
 OY 181 SFSK 184  
 DB 383 SFSK 386

## RESULT 7

O35206 PRELIMINARY; PRT; 1367 AA.  
 AC 035206;

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
 DE PROCOLLAGEN, TYPE XV (TYPE XV COLLAGEN).

```

CN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains".
RL Genomics 45:31-41(1997).
DR EMBL; AF011450; AAC53387.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88449; COL15a1.
DR InterPro; IPR000087; -.
DR InterPro; IPR003129; -.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; LamG; 1.
SQ SEQUENCE 1367 AA; 140524 MW; AA83A1254AF3AEEC CRC64;

Query Match 59.3%; Score 574; DB 11; Length 1367;
Best Local Similarity 62.4%; Pred. No. 4,4e-49;
Matches 111; Conservative 22; Mismatches 41; Indels 4; Gaps 1;

QY 7 OPIVHLVLTPTLPGSGMRGIRGADFCQFOARAVGLSGTFRALSSRLDPLSYVRADR 66
DB 1194 RPYLHLVLTPTPAAGDIR---ADFCFOARAGLSTFRALSSRLDPLSYVRADR 1249
QY 67 GSVPIVNLKDEVLPSPWDSLFSGSGQLQPGARIFSPDGRDVLRRHPAPKQSVWHSNPS 126
DB 1250 FGLPIVNLKGVLFNNWDSLFSGDGFNTHTPIYSPDGRDVTDPSPQKVVWHSNPH 1309
QY 127 GRRLMESYCEWTRETTGATGQASSLSGRLLDEKAAKSNYSIVLCIENSFMFTSPK 184
DB 1310 GVRIVDKYCEAWRTTDMAYVGFASPLSTGKILDKKAYSCANRLIVLCIENSFMFTDRK 1367

RESULT 8
Q9EOD9 PRELIMINARY; PRT; 1367 AA.
AC Q9EOD9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=129/SV;
RX MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains".
RL Genomics 45:31-41(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=129/SV;
RX PubMed=11068203;
RA Eklund L., Muona A., Liefard J., Pihlajaniemi T.;

```

```

RT "Structure of the mouse type XV collagen gene, Col15a1, comparison
RT with the human COL15A1 gene and functional analysis of the promoters
RT of both genes.";
RL Matrix Biol. 19:489-500(2000).
DR EMBL; AF261131; AAG27545.1; -.
DR EMBL; AF261109; AAG27545.1; JOINED.
DR EMBL; AF261110; AAG27545.1; JOINED.
DR EMBL; AF261111; AAG27545.1; JOINED.
DR EMBL; AF261112; AAG27545.1; JOINED.
DR EMBL; AF261113; AAG27545.1; JOINED.
DR EMBL; AF261114; AAG27545.1; JOINED.
DR EMBL; AF261115; AAG27545.1; JOINED.
DR EMBL; AF261116; AAG27545.1; JOINED.
DR EMBL; AF261117; AAG27545.1; JOINED.
DR EMBL; AF261118; AAG27545.1; JOINED.
DR EMBL; AF261119; AAG27545.1; JOINED.
DR EMBL; AF261120; AAG27545.1; JOINED.
DR EMBL; AF261121; AAG27545.1; JOINED.
DR EMBL; AF261122; AAG27545.1; JOINED.
DR EMBL; AF261123; AAG27545.1; JOINED.
DR EMBL; AF261124; AAG27545.1; JOINED.
DR EMBL; AF261125; AAG27545.1; JOINED.
DR EMBL; AF261126; AAG27545.1; JOINED.
DR EMBL; AF261127; AAG27545.1; JOINED.
DR EMBL; AF261128; AAG27545.1; JOINED.
DR EMBL; AF261129; AAG27545.1; JOINED.
DR EMBL; AF261130; AAG27545.1; JOINED.
SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;

Query Match 59.2%; Score 573; DB 11; Length 1367;
Best Local Similarity 62.4%; Pred. No. 5,5e-49;
Matches 111; Conservative 22; Mismatches 41; Indels 4; Gaps 1;

QY 7 OPIVHLVLTPTLPGSGMRGIRGADFCQFOARAVGLSGTFRALSSRLDPLSYVRADR 66
DB 1194 RPYLHLVLTPTPAAGDIR---ADFCFOARAGLSTFRALSSRLDPLSYVRADR 1249
QY 67 GSVPIVNLKDEVLPSPWDSLFSGSGQLQPGARIFSPDGRDVLRRHPAPKQSVWHSNPS 126
DB 1250 FGLPIVNLKGVLFNNWDSLFSGDGFNTHTPIYSPDGRDVTDPSPQKVVWHSNPH 1309
QY 127 GRRLMESYCEWTRETTGATGQASSLSGRLLDEKAAKSNYSIVLCIENSFMFTSPK 184
DB 1310 GVRIVDKYCEAWRTTDMAYVGFASPLSTGKILDKKAYSCANRLIVLCIENSFMFTDRK 1367

RESULT 9
Q9Y4W4 PRELIMINARY; PRT; 1388 AA.
AC Q9Y4W4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94148920; PubMed=8106446;
RA Kivirikko S., Heilmaki P., Rehn M., Honkanen N., Myers J.C.,
RA Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of human type XV collagen and
RT exon-intron organization in the 3' region of the corresponding gene.";
RL J. Biol. Chem. 269:4773-4779(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98316357; PubMed=9651385;
RA Hagg P.M., Muona A., Liefard J., Kivirikko S., Pihlajaniemi T.;

```

RL J. Biol. Chem. 273:17824-17831 (1998)

QY 7 QPVLHLVALNTPLSGGMGRIGADFOCFOARAVGLSGTFRALSSRIQDIYSTVBRAB 66

$$R_N [1]$$

RX MEDLINE=20196006; PubMed=10731132;

8 PYLHLYAI<sup>NTPI</sup>SGGMBSTBGAD<sup>EOCF</sup>EO<sup>QAR</sup>AVGI<sup>SGTE</sup>PA<sup>IS</sup>SP<sup>ION</sup>IY<sup>STV</sup>BA<sup>NBC</sup> 57

DT 01-JAN-1998 (TREMBlEt. 05, Created





---

	RESULT	1
AA849380	ID	AA849380 standard; protein; 164 AA.
AC	XX	AA849380;
AD	XX	02-MAR-2001 (first entry)
DE	XX	Murine endostatin SEQ ID NO: 4.
KW	XX	Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
KW	XX	cancer; inflammation; angiogenesis-dependent disease.
OS	XX	Mus musculus.
PN	XX	WC2000067771-A1.
PD	XX	16-NOV-2000.
PF	XX	02-MAY-2000; 2000MO-US12063.
PR	XX	06-MAY-1999; 99US-0132907.
PR	XX	14-JUL-1999; 99US-0353333.
PA	XX	(BURN-) BURHAM INST.
PI	XX	Vuori K;
DR	XX	WPI; 2001-040937/05.
DR	XX	N-PDB; AAC88290.
PT	XX	Endostatin peptide comprising at least four endostatin amino acid
PT	XX	residues are e.g. angiogenesis inhibitors for treating cancer and
PT	XX	diabetic retinopathy -

XX Disclosure; Fig 1; 146pp; English.  
PS

The present invention provides endostatin peptides which can be used in the modulation of angiogenesis. This is useful in the treatment of cancers, inflammation, rheumatoid arthritis, chronic articular rheumatism, psoriasis, disorders associated with inappropriate invasion of vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy of prematurity, macular degeneration, corneal graft rejection, retrolental fibroplasia, rubecosis, capillary proliferation in atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque neovascularisation, telangiectasia, hemophilic joints and wound granulation. In addition, the peptides can be used as birth control agents.

**50**      **Sequence**      **184 AA;**

Query Match	100.0%	Score 968;	DB 22;	Length 184;
Best Local Similarity	100.0%	Pred. No. 3.2e-108;		
Matches 184; Conservative	0;	Mismatches	0;	Gaps 0;

Qy 1 HTHDQFQVHLVALNPPLSGGMRGIRKADFOCFQARAVGIGSTFRFLSSRKDDLYSI 600  
|||||  
Db 1 hthdqpvhhlvalncplsgmqrllrqadfcfcqgaravglsgtffraflssrlqdlysi 600

```

0y      61 VRRADKGSVPIVNIKKDEVLSBWSDSLSSGSGGQLQPGARISFDGRDYLRRHPAWQKSVW 120
        |||||||
0b      61 VRRADKGSVPIVNIKKDEVLSBWSDSLSSGSGGQLQPGARISFDGRDYLRRHPAWQKSVW 120

```

121 HGSDPSGRRLMESCYETWRTETTGATGQASSLLSGRLLEQKAAASCHNSYIVLCIENSFMT 180

OY	181	SFSK	184
Db	181	SFSK	184

RESULT	2
AAB28398	
ID	AAB28398 standard; Protein; 191 AA.

AC	AAB28398;
XX	
DT	19-FEB-2001 (first entry)

DE Murline endostatatn.

KM Murine; endostatin; cytostatic; antiproliferative;  
 KM vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
 KM cancer; vascularised solid tumour.

Mus sp.

PN WO200064946-A2.

PD 02-NOV-2000.

PF 28-APR-2000; 2000WO-US11367.

PR 28-APR-1999; 99US-0131432.

PA (TEXA ) UNIV TEXAS SYSTEM.

PI Thorpe PE, Brekken RA;

DR WPI; 2000-687317/67.

DR N-PSDB; AAC67777.

PT Immunogenic composition for the treatment and diagnosis of cancer  
PT comprises an anti-VEGF (vascular endothelial growth factor) antibody

PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -  
XX  
PS Example 10; Page 290-291; 298pp; English.

CC The present invention relates to anti-Vascular Endothelial Growth Factor  
CC (VEGF) antibodies that bind to the same epitope as the monoclonal  
CC antibody ATCC P1A 1593 and which significantly inhibit VEGF binding to  
CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF  
CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin  
CC may be conjugated onto the anti-VEGF antibodies of the present invention.  
CC The anti-VEGF antibodies of the present invention are useful for the  
CC treatment and diagnosis of cancer especially vascularised solid tumours

Sequence 191 AA;

Query Match	100.0%	Score 968	DB 21	Length 191
Best Local Similarity	100.0%	Pred. No. 3.3e-108		
Matches 184	Conservative 0	Mismatches 0	Indels 0	Gaps 0

**Qy** 1 HTHDGFQPVLHVALNPLSGMGIRGADFQCFOARAVGLSTFRFLSSRKQDLYSI 60  
**Db** 8 hthdgtfpvhlvalncplsgmgilrgadfqcfqgaravglstftafllssrlqlysl 67			

QY 61 VRRADRGSPVIVNLKDEVLSPSMDSLFSSGSGQLPGARLFSFDGRDVLRRHPAMPQKSVW 120

DB 68 VRRDGRSPVIVNLKDEVLSPSMDSLFSSGSGQLPGARLFSFDGRDVLRRHPAMPQKSVW 127

QY 121 HGSDDSGRLMESYCEFWRTETTGATGCASSLGRLLEPKAASCHNSYIVICIEFSMT 1800  
|||||  
Db 128 hsdhdsrrrimsavcvtvrfelrtatnaaa|||vrr||leekaschshv|v|clensfnt 1872

QY	181	SFSK	184
Db	188	SFSK	191

RESULT	3
AAV08689	
ID	AAV08689 standard; Protein; 184 AA.

AC AAY08689;

DT 10-AUG-1999 (first entry)

DE Murine endostatin protein fragment.

KM Plasmalogen; murine; angiotensin; endostatin; gene therapy; vector;  
KM anti-angiogenic; attenuation; cytosolic; anti-diabetic; ophthalmology;  
KM tumour growth; solid tumour; diabetic retinopathy; retina.

OS Mus sp.

PN W09926480-A1

PD 03-JUN-1999

PF 20-NOV-1998; 98WO-US24950.

PR 20-NOV-1997; 97US-0975424.

PA (GENE-) GENETIX PHARM INC.  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.

PI Bachelot T, Leboulch P, Pawluk RJ;

DR WPI; 1999-357696/30.

DR N-PSDB; AAX77715.

PT Ant1-angiotensin gene therapy vectors

PS Disclosure; Fig 6; 83pp; English

CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiostatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells *ex vivo* and then administered to the patient.  
XX  
SQ Sequence 184 AA:

Query Match 99.7%; Score 965; DB 20; Length 184;  
Best Local Similarity 99.5%; Pred. No. 7.3e-108;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHQDFQPVHLALNTPLSGMRGIRGADFOCFQOARAVGLSGTPRAFLSSRLDLYSI 60  
DB 1 hthqdfqpvhlalntplsgmrgirgadfgcfqgaravglsgtrflssrlqlysl 60  
OY 61 VRRADRGSPVPIYNLKDENVLSFSGSQGLQPGARIFSGDGVLRHPAMPQKSYW 120  
DB 61 vrradrgsvpiynlkdevlspswdsifsgsqgvgpgarifsgdgvlrhpmpqksw 120  
OY 121 HGSDPSGRRLMESYCTWRTETTGATGOASSLSLGRLEOKAASCHNSYIVLCIENSFMT 180  
DB 121 hgsdpsgrrlmesyctwrtetgtgqassllsgrlleqkaaschmsyivlcienfnt 180  
OY 181 SFSK 184  
DB 181 sfsk 184

RESULT 4  
AA70258

ID AAY70258 standard; Protein: 184 AA.

XX AAY70258;

DE 06-JUN-2000 (first entry)

XX Murine angiogenesis inhibitor, endostatin.

KW Murine; immunoglobulin Fc fragment; endostatin; immunofusin;  
KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;  
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;  
KW vasotrophic; vulnery; treatment; antiarteriosclerosis; tumour;  
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
KW ocular angiogenesis; diabetic retinopathy; macular degeneration;  
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
KW wound granulation; keloid scar; gene therapy.

XX Mus musculus.

OS WO200011033-A2.

PN 02-MAR-2000.

XX 25-AUG-1999; 99WO-US19329.

XX 25-AUG-1998; 98US-0097883.

XX (LEXI-) LEXINGEN PHARM CORP.

XX Le K, Li Y, Gillies SD;

XX WPI, 2000-237616/20.

DR N-PSDB; AA251299.

PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin  
PT FC region, useful for treating conditions mediated by angiogenesis,  
PT such as rheumatoid arthritis, tumors and macular degeneration -  
XX  
XX Example 5; Pages 48-49; 68pp; English.

CC The patent discloses a DNA molecule encoding a fusion protein comprising  
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment  
CC having angiostatin activity, a collagen XVIII fragment having endostatin  
CC activity, or combinations of them. The fusion protein (immunofusin) is  
CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
CC by angiogenesis. Conditions that may be treated include solid tumours,  
CC blood born tumours, tumour metastasis, benign tumours including  
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
CC fibroplasia, rubecosis and Osler-Weber syndrome; myocardial angiogenesis,  
CC plaque neovascularisation, telangiectasia, haemophilic joints,  
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of  
CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and  
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
CC in gene therapy. The present sequence is a murine  
CC endostatin used in the construction of immunofusin containing murine  
CC immunoglobulin Fc fragment.

SQ Sequence 184 AA:

Query Match 99.7%; Score 965; DB 21; Length 184;  
Best Local Similarity 99.5%; Pred. No. 7.3e-108;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHQDFQPVHLALNTPLSGMRGIRGADFOCFQOARAVGLSGTPRAFLSSRLDLYSI 60  
DB 1 hthqdfqpvhlalntplsgmrgirgadfgcfqgaravglsgtrflssrlqlysl 60  
OY 61 VRRADRGSPVPIYNLKDENVLSFSGSQGLQPGARIFSGDGVLRHPAMPQKSYW 120  
DB 61 vrradrgsvpiynlkdevlspswdsifsgsqgvgpgarifsgdgvlrhpmpqksw 120  
OY 121 HGSDPSGRRLMESYCTWRTETTGATGOASSLSLGRLEOKAASCHNSYIVLCIENSFMT 180  
DB 121 hgsdpsgrrlmesyctwrtetgtgqassllsgrlleqkaaschmsyivlcienfnt 180  
OY 181 SFSK 184  
DB 181 sfsk 184

RESULT 5  
AAB71930

ID AAB71930 standard; Protein: 207 AA.

XX AAB71930;

DE 10-MAY-2001 (first entry)

XX Murine endostatin attached to Ig-kappa leader sequence.

KW Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;  
KW antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;  
KW adenoviral vector; diabetic retinopathy; cardiovascular disease;  
KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;  
KW lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.

XX Mus sp.

XX WO200112830-A1.

XX 22-FEB-2001.

PF 11-AUG-2000; 2000MO-BE07865.  
 XX  
 PR 13-AUG-1999; 99US-0373938.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 XX  
 PI Hallenbeck PL, Chen CT;  
 XX  
 DR WPI: 2001-202871/20.  
 DR N-PSDB; AAF60336.  
 PT Adenoviral vector for treating tumors and disorders associated with  
 PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA  
 PT sequence encoding an angiogenic inhibitor, particularly endostatin  
 XX  
 PS Example 1; Fig 1B; 59pp; English.  
 XX  
 CC The nucleotide sequence encoding this protein was used in the  
 CC construction of an adenoviral vector which includes a DNA sequence  
 CC encoding endostatin. The adenoviral vector is useful for expressing  
 CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is  
 CC useful for treating other diseases and disorders associated with  
 CC angiogenesis, such as neovascular diseases of the eye, including diabetic  
 CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral  
 CC edema and intravascular coagulopathy (Kasabach-Merrill syndrome). The  
 CC vector inhibits, prevents or destroys the growth of tumors by  
 CC preventing the formation of blood vessels in tumors, such as lymphoma  
 CC and leukaemia.  
 XX  
 SQ Sequence 207 AA:  
 XX  
 Query Match 99.7%; Score 965; DB 22; Length 207;  
 Best Local Similarity 99.5%; Pred. No. 8.6e-108;  
 Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HTHDDFQPVHLVALNTPLSGKNGIRGADFCFOQARAVGLSGTFRAFLLSRKQDLYSI 60  
 DB htdqdfqpvhlvalntrplsgmrgirgadfcfgqaravglsgtfrfaflsrldlysl 83  
 QY 61 VRRADKRSVPIVNLKDEVLSFSDSLFSGGQGLQPCARIFSPDGRVYLHRPMPQKSVW 120  
 DB vrradkrsvpiivnlkdevlspswdsltsgsgqglqpcarlfspdgrrvylhrmpqksw 143  
 QY 121 HGSDFSGRRLMESYCFWRETGTGATGQASSLSGRLLLEQKAASCHNSYIVLCIENSFMT 180  
 DB hgsdfsgrrlmesyctwretgtgatgqassllsgrllleqkaaschnsyivlcienfnt 203  
 QY 181 SFSK 184  
 DB 204 sfsk 207  
 XX  
 RESULT 6  
 ID AAY08691 standard; Protein: 218 AA.  
 XX  
 AC AAY08691;  
 XX  
 DT 10-AUG-1999 (first entry)  
 XX  
 DE Murine gene therapy peptide construct SP-Flag-Endo.  
 XX  
 KW Plasmidogen: murine; angiotatin; endostatin; gene therapy; vector;  
 KW anti-angiogenic; attenuation; cytosstatic; anti-diabetic; ophthalmology;  
 KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.  
 XX  
 OS Mus sp.  
 OS Synthetic.  
 OS  
 PN W09926480-A1.  
 XX

PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98MO-US24950.  
 XX  
 PR 20-NOV-1997; 97US-0975424.  
 XX  
 PA (GENE-) GENETIX PHARM INC.  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Bachelot T, Leboulch P, Pawliuk RJ;  
 XX  
 DR WPI: 1999-357696/30.  
 DR N-PSDB; AAX77717.  
 PT Anti-angiogenic gene therapy vectors  
 XX  
 PS Example 1; Page 69; 83pp; English.  
 XX  
 CC This invention describes a novel viral gene therapy vector comprising a  
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
 CC from human or murine angiotatin, human or murine endostatin and  
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
 CC sufficiently attenuated for use in human gene therapy. The products of  
 CC the invention have anti-angiogenic, cytosstatic, anti-diabetic and  
 CC ophthalmological activity. The vector is used in gene therapy for  
 CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
 CC expresses an anti-angiogenic polypeptide. An additional use comprises  
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
 CC inhibits angiogenesis in the vicinity of the retina. The vector is  
 CC administered to cells ex vivo and then administered to the patient.  
 XX  
 SQ Sequence 218 AA:  
 XX  
 Query Match 99.7%; Score 965; DB 20; Length 218;  
 Best Local Similarity 99.5%; Pred. No. 9.3e-108;  
 Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HTHDDFQPVHLVALNTPLSGKNGIRGADFCFOQARAVGLSGTFRAFLLSRKQDLYSI 60  
 DB htdqdfqpvhlvalntrplsgmrgirgadfcfgqaravglsgtfrfaflsrldlysl 94  
 QY 61 VRRADKRSVPIVNLKDEVLSFSDSLFSGGQGLQPCARIFSPDGRVYLHRPMPQKSVW 120  
 DB vrradkrsvpiivnlkdevlspswdsltsgsgqglqpcarlfspdgrrvylhrmpqksw 154  
 QY 121 HGSDFSGRRLMESYCFWRETGTGATGQASSLSGRLLLEQKAASCHNSYIVLCIENSFMT 180  
 DB hgsdfsgrrlmesyctwretgtgatgqassllsgrllleqkaaschnsyivlcienfnt 214  
 QY 181 SFSK 184  
 DB 215 sfsk 218  
 XX  
 RESULT 7  
 ID AAY08692 standard; Protein: 580 AA.  
 XX  
 AC AAY08692;  
 XX  
 DT 10-AUG-1999 (first entry)  
 XX  
 DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.  
 XX  
 KW Plasmidogen: murine; angiotatin; endostatin; gene therapy; vector;  
 KW anti-angiogenic; attenuation; cytosstatic; anti-diabetic; ophthalmology;  
 KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.  
 XX  
 OS Mus sp.  
 OS Synthetic.  
 OS  
 PN W09926480-A1.  
 XX

XX 03-JUN-1999.  
PD 20-NOV-1998; 98MO-US24950.  
XX 20-NOV-1997; 97US-0975424.  
PR (GENE-) GENETIX PHARM INC.  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX Bachelot T, Leboulch P, Pawluk RJ;  
XX WPI: 1999-357696/30.  
DR N-PSDB; AAX77718.  
XX Anti-angiogenic gene therapy vectors  
PT Example 1; Page 72-74; 83pp; English.  
PS This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiotatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.  
XX  
SQ Sequence 580 AA:  
  
Query Match 99.7%; Score 965; DB 20; Length 580;  
Best Local Similarity 99.5%; Pred. No. 4e-107;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HTHDFQPVLRHVALNTPLSGMRGIRGADFOCFQOARAVGLSGTRAPLSSRLDPLYSI 60  
DB 397 hthdftqpvlrhvalntrplsgmrgirgadfcqfqtaravglsgtratlssrlqdlysi 456  
QY 61 VRADRGSVPIVNLDEVLSFSSGQQLPGARIFSPDGRDVLRRHPMPQKSYW 120  
DB 457 vrradrgsvpiivnldevlsfssgqqlpgarifspdgrrdvlrrhmpqksw 516  
QY 121 HGSDEPSGRRLMESYCTWRTTGATGCASSLSSGRLLEOKAKASCHNSIYVICIENSFMT 180  
DB 517 hgsdpsgrrlmesyctwrttgcagassllsgrllleqkaaschnsyiviciensfnt 576  
QY 181 SFSK 184  
DB 577 sfsk 580  
  
RESULT 8  
AAW26328  
ID AAW26328 standard; Protein; 1288 AA.  
AC AAW26328;  
XX 19-NOV-1997 (first entry)  
DE Mouse alpha-1 collagen (XVIII).  
XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration.  
XX Mms musculus.  
OS  
XX  
XX Key Location/Qualifiers  
FH Peptide 303..308  
FT /label- GXYGX'Y'\_motif

FT Peptide 309..314  
FT /label- GXYGX'Y'\_motif  
FT Peptide 315..320  
FT /label- GXYGX'Y'\_motif  
FT Peptide 321..326  
FT /label- GXYGX'Y'\_motif  
FT Peptide 337..342  
FT /label- GXYGX'Y'\_motif  
FT Peptide 343..348  
FT /label- GXYGX'Y'\_motif  
FT Peptide 349..354  
FT /label- GXYGX'Y'\_motif  
FT Peptide 355..360  
FT /label- GXYGX'Y'\_motif  
FT Peptide 361..366  
FT /label- GXYGX'Y'\_motif  
FT Peptide 367..372  
FT /label- GXYGX'Y'\_motif  
FT Peptide 373..378  
FT /label- GXYGX'Y'\_motif  
FT Peptide 379..384  
FT /label- GXYGX'Y'\_motif  
FT Peptide 385..390  
FT /label- GXYGX'Y'\_motif  
FT Peptide 396..401  
FT /label- GXYGX'Y'\_motif  
FT Peptide 402..407  
FT /label- GXYGX'Y'\_motif  
FT Peptide 435..440  
FT /label- GXYGX'Y'\_motif  
FT Peptide 441..446  
FT /label- GXYGX'Y'\_motif  
FT Peptide 447..452  
FT /label- GXYGX'Y'\_motif  
FT Peptide 453..458  
FT /label- GXYGX'Y'\_motif  
FT Peptide 459..464  
FT /label- GXYGX'Y'\_motif  
FT Peptide 470..475  
FT /label- GXYGX'Y'\_motif  
FT Peptide 476..481  
FT /label- GXYGX'Y'\_motif  
FT Peptide 482..487  
FT /label- GXYGX'Y'\_motif  
FT Peptide 488..493  
FT /label- GXYGX'Y'\_motif  
FT Peptide 494..499  
FT /label- GXYGX'Y'\_motif  
FT Peptide 500..505  
FT /label- GXYGX'Y'\_motif  
FT Peptide 506..511  
FT /label- GXYGX'Y'\_motif  
FT Peptide 512..517  
FT /label- GXYGX'Y'\_motif  
FT Peptide 518..523  
FT /label- GXYGX'Y'\_motif  
FT Peptide 524..529  
FT /label- GXYGX'Y'\_motif  
FT Peptide 530..535  
FT /label- GXYGX'Y'\_motif  
FT Peptide 536..541  
FT /label- GXYGX'Y'\_motif  
FT Peptide 542..547  
FT /label- GXYGX'Y'\_motif  
FT Peptide 548..553  
FT /label- GXYGX'Y'\_motif  
FT Peptide 580..585  
FT /label- GXYGX'Y'\_motif  
FT Peptide 586..591  
FT /label- GXYGX'Y'\_motif  
FT Peptide 592..597  
FT /label- GXYGX'Y'\_motif  
FT Peptide 598..603



OSler-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;  
KM plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;  
KM dialysis graft vascular access stenosis; renal cancer; therapy.  
XX  
OS Mus sp.  
XX  
PN MO9929855-A1.  
XX  
PD 17-JUN-1999.  
XX  
PF 08-DEC-1998: 98WO-US26057.  
XX  
PR 16-NOV-1998: 98US-0108536.  
PR 08-DEC-1997: 97US-0067888.  
PR 22-APR-1998: 98US-0082663.  
XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
PI Sukhame VP.  
XX  
DR WPI: 1999-385604/32.  
DR N-PSDB; AAX79949.  
XX  
PT Mutant endostatin having anti-angiogenic activity  
XX  
PS Claim 31; Fig 2; 105pp; English.  
XX  
CC This sequence is the mouse endostatin. The invention relates to a  
CC the mutant endostatin (EM), which has anti-angiogenic activity, and is  
CC designated EM1. Compositions comprising EM1 or fusion proteins comprising  
CC EM1, are useful for treating diseases characterised by angiogenic  
CC activity, such as angiogenesis-dependent cancers, benign tumours,  
CC rheumatoid arthritis, psoriasis, ocular angiogenesis, OSler-Webber  
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,  
CC telangiectasia, haemophilic joints, angiofibroma, wound granulation,  
CC intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,  
CC cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular  
CC access stenosis, contraception and obesity. In particular, the diseases  
CC treatable by EM1 comprise cancer, especially renal cancer. The methods  
CC provide a means for introducing EM1 into mammalian cells via gene  
CC therapy, for production of EM1 via recombinant means, as well as  
CC recombinant production of the EM1 protein. EM1 performs as well or better  
CC than whole endostatin. Use of EM1 is advantageous for treatment of  
CC angiogenic diseases in that increasingly smaller peptides are more potent  
CC on a weight basis, and may be able to better penetrate tissues.  
CC  
XX  
SQ Sequence 184 AA:  
  
Query Match 99.5%; Score 963; DB 20; Length 184;  
Best Local Similarity 99.5%; Pred. No. 1.3e-107;  
Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 HTHOFOPLVHLVANTPLSGMGRIGRAGDPOCARAVGUSGFRFLSRLODLYSI 60  
1 hthgfgpvlhvalncplsggmrglrgadlqctqgaravgsiftrflssrlldlysi 60  
Db 1 hthgfgpvlhvalncplsggmrglrgadlqctqgaravgsiftrflssrlldlysi 60  
OY 61 VRRADRGSVPIVNLKDEVLSPSMDSLFSGSOGLOPGARIFSPDRDVLRRHPAMPQKSVW 120  
61 vrradrgsvpivnlkdevlspmdslfsgsgqlpgarilfsfgrdvlrrhpampqksw 120  
Db 61 vrradrgsvpivnlkdevlspmdslfsgsgqlpgarilfsfgrdvlrrhpampqksw 120  
OY 121 HGSDDSGRRIMESYCEWRTETTGATGOASSLSGRLLERKASCHNSYIVLCIENSFMT 180  
121 hgsddsgrrimesycewrtetttgatgoasslsgrllerkaschnsyivlcienfnt 180  
Db 121 hgsddsgrrimesycewrtetttgatgoasslsgrllerkaschnsyivlcienfnt 180  
OY 181 SFSK 184  
181 sfsk 184  
Db 181 sfsk 184

RESULT 10  
AAW92297 X  
ID AAW92297 standard; peptide: 1288 AA.

XX  
AC AAW92297;  
XX  
DT 28-APR-1999 (first entry)  
XX  
DE Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.  
XX  
KM Human; type XVIII collagen; liver disease; cirrhosis; detection;  
KM hepatocellular carcinoma; diagnosis.  
XX  
OS Mus sp.  
XX  
PN MO9856399-A1.  
XX  
PD 17-DEC-1998.  
XX  
PF 12-JUN-1998: 98WO-US12327.  
XX  
PR 12-JUN-1997: 97US-0049369.  
XX  
PA (FIB-) ACAD FINLAND.  
PA (FIB-) FIBROGEN INC.  
PA (INRM) INST NAT SANTE & RECH MEDICALE.  
XX  
PI Clement B, Pihlajaniemi T, Rehn M:  
XX  
DR WPI: 1999-070292/06.  
XX  
PT Diagnosis and monitoring of liver disease by measuring collagen type  
PT XVIII levels - with elevated levels indicative of disease,  
XX especially cirrhosis or hepatocellular carcinoma  
XX  
PS Example 6; Fig 8; 56pp; English.  
XX  
CC A method has been developed for the detecting liver disease. The method  
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific  
CC for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen  
CC complex (c) formed as indicator of the amount of Col18 present; (c)  
CC similar analysis of a non-diseased control; and (d) comparing the  
CC amounts of Col18 in the two samples to detect presence or progression of  
CC disease. Elevated levels of Col18 are: (i) indicative of disease,  
CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,  
CC specifically hepatocellular carcinoma (there is a relationship between  
CC Col18 mRNA levels and tumour size and necrosis, and survival times are  
CC significantly higher in patients with higher Col18 levels). The method  
CC provides non-invasive, early and accurate diagnosis of liver disease.  
CC The present sequence represents the sequence common to mouse alpha-1  
CC (XVIII) collagen chain from the present invention.  
XX  
SQ Sequence 1288 AA:  
  
Query Match 99.2%; Score 960; DB 20; Length 1288;  
Best Local Similarity 99.5%; Pred. No. 5.2e-106;  
Matches 182; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 HTHOFOPLVHLVANTPLSGMGRIGRAGDPOCARAVGUSGFRFLSRLODLYSI 60  
1 hthgfgpvlhvalncplsggmrglrgadlqctqgaravgsiftrflssrlldlysi 1165  
Db 1166 hthgfgpvlhvalncplsggmrglrgadlqctqgaravgsiftrflssrlldlysi 1165  
OY 61 VRRADRGSVPIVNLKDEVLSPSMDSLFSGSOGLOPGARIFSPDRDVLRRHPAMPQKSVW 120  
61 vrradrgsvpivnlkdevlspmdslfsgsgqlpgarilfsfgrdvlrrhpampqksw 1225  
Db 1166 vrradrgsvpivnlkdevlspmdslfsgsgqlpgarilfsfgrdvlrrhpampqksw 1225  
OY 121 HGSDDSGRRIMESYCEWRTETTGATGOASSLSGRLLERKASCHNSYIVLCIENSFMT 180  
121 hgsddsgrrimesycewrtetttgatgoasslsgrllerkaschnsyivlcienfnt 1285  
Db 1226 hgsddsgrrimesycewrtetttgatgoasslsgrllerkaschnsyivlcienfnt 1285  
OY 181 SFS 183  
181 sfs 183  
Db 181 sfs 183

RESULT 11  
AAV25114  
ID AAV25114 standard; Protein; 684 AA.  
XX  
AC AAV25114;  
XX  
DT 25-AUG-1999 (first entry)  
XX  
DE Mouse alpha1 (XVIII) collagen protein.  
XX  
KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;  
KW anti-angiogenic; heparin binding domain; receptor binding domain; mimic;  
KW alpha-helix A domain; carbohydrate recognition domain; CRD domain;  
KW treatment; angiogenesis; tumour; murine.  
XX  
OS Mus sp.  
XX  
PN MO9931616-A1.\*  
XX  
PD 24-JUN-1999.  
XX  
PE 16-DEC-1998; 98MO-US26783.  
XX  
PR 16-DEC-1997; 97US-0069727.  
XX  
PA (HARD ) HARVARD COLLEGE.  
XX  
PI Hohenester E, Olsen BR, Sasaki T, Timpl R;  
XX  
DR WPI; 1999-395243/33.  
XX  
PT Identifying mimetics of mammalian endostatin  
XX  
PS Disclosure: Fig 5A-C; 75pp; English.  
XX  
CC This invention describes a novel method for identifying mimetics of  
CC mammalian endostatin. The method comprises identifying a compound  
CC having atomic coordinates with non-trivial similarity to selected  
CC coordinates of atoms of a mammalian endostatin involves (a) providing  
CC a library of atomic coordinates of compounds in a library of candidate  
CC compounds, (b) comparing the library of atomic coordinates to the  
CC selected coordinates of a mammalian endostatin and (c) selecting from the  
CC library at least one candidate compound on the basis of selection  
CC criteria which include similarities between the atomic coordinates of the  
CC selected candidate compound and the atomic coordinates of the mammalian  
CC endostatin. The invention also describes the use of an anti-angiogenic  
CC fragment of endostatin comprising a domain selected from a heparin  
CC binding domain, a receptor binding domain, and exposed on alpha-helix A  
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods  
CC can be used for designing and selecting endostatin mimics. The compounds  
CC identified can be used for treating undesired angiogenesis, e.g. tumours.  
CC This sequence represents mouse alpha1(XVIII) collagen which is used in  
CC the description of the method.  
XX  
XX  
Sequence 684 AA:  
50  
Query Match 98.18; Score 950; DB 20; Length 684;  
Best Local Similarity 99.48; Pred. No. 3.2e-105;  
Matches 180; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 181 S 181  
DB 682 S 682  
RESULT 12  
AAV06197  
ID AAV06197 standard; Protein; 185 AA.  
XX  
AC AAV06197;  
XX  
DT 16-AUG-1999 (first entry)  
XX  
DE Anti-angiogenic endostatin peptide.  
XX  
KW Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human;  
KW endostatin; melanoma; lung cancer; colon cancer; brain cancer;  
KW breast cancer.  
XX  
OS Homo sapiens.  
XX  
PN  
XX  
EH Key Location/Qualifiers  
FT Misc-difference 36 /note= "encoded by CAG"  
FT Misc-difference 37 /note= "encoded by CAA"  
FT Misc-difference 39 /note= "deduced sequence from nucleotide sequence  
FT /note= "has an Ala residue between residues 39 and  
FT 40 of this sequence"  
FT Misc-difference 76 /note= "encoded by AAG"  
FT Misc-difference 118 /note= "encoded by AAG"  
FT Misc-difference 162 /note= "encoded by AAA"  
FT Misc-difference 168 /note= "encoded by AAC"  
FT Misc-difference 185 /note= "encoded by AAA"  
FT  
XX  
PN EP921193-A1.  
XX  
PD 09-JUN-1999.  
XX  
PE 07-JAN-1998; 98EP-0100135.  
XX  
PR 05-DEC-1997; 97US-0985526.  
XX  
PA (MIXS/) MIXSON A J.  
XX  
PI Mixson AJ;  
XX  
DR WPI; 1999-315406/27.  
DR N-PSDB; AAX58740.  
XX  
PT  
XX  
PS Disclosure: Page 38; 46pp; English.  
XX  
CC The present sequence represents an anti-angiogenic endostatin  
CC peptide. The invention provides a carrier:DNA complex that comprises  
CC DNA (see AAX58725-42) encoding an anti-angiogenic protein or peptide,  
CC such as the present sequence, the complex being deliverable to  
CC the site of a tumour in vivo, and additionally comprises regulatory  
CC elements for expressing the anti-angiogenic DNA in a tumour or  
CC tumour vasculature. The complex may also include DNA encoding a  
CC tumour suppressor protein, especially p53. The carrier is a  
CC liposome, cationic polymer, micelle, microsphere, virus, viral  
CC component, or a combination of these, and administration is by  
CC intravenous or intratumoral injection. The complexes are useful in  
CC gene therapy for inhibition of tumour growth. The types of tumors  
CC which may be treated include solid tumors such as melanomas and

CC tumors in the lung, colon, brain and breast.  
 XX Sequence 185 AA;

Query Match 94.6%; Score 916; DB 20; Length 185;  
 Best Local Similarity 95.1%; Pred. No. 5,7e-102;  
 Matches 176; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

QY 1 HTHQDFOPVHLVNLTPLSGSGMRIGADPCQARAVGLSGFRFLSSRLQDLKST 60  
 Db 2 hthqdfqpyhlvnltpslsgsmrgirgadfqrtnar-vqlsgrlraflssrlqdlkyl 60  
 QY 61 VRADRGSPVIV-NLKDEVLSFSDSLFSGSQGLQPCARIFSPGDRVLRHPAPQKSV 119  
 Db 61 vrradrgspvivrnlkdevlfspswslfsgsqglqpcarifsfqdrvrlrhpapqrsy 120  
 QY 120 WHGSDPSGRRLMESYCEWRTETTGATGQASSLSGRLLBOKRASCCHNSYIVLCIENSFM 179  
 Db 121 whgsdpsgrllmesycewrtetgtgassllsgrlllegraaschdsyivlcienfsm 180  
 QY 180 TSFSK 184  
 Db 181 tsfsr 185

RESULT 13  
 AAY70265  
 ID AAY70265 standard; Protein; 184 AA.  
 AC AAY70265;

DT 06-JUN-2000 (first entry)  
 XX Canine angiogenesis inhibitor, endostatin.  
 DE Canine angiogenesis inhibitor, endostatin.

KW Canine: immunoglobulin Fc fragment; endostatin; immunofusin;  
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritis;  
 KW antiproliferative; antidiabetic; ophthalmological; immunosuppressant;  
 KW vasotrophic; vulnary; treatment; antiarteriosclerosis; tumor;  
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
 KW wound granulation; keloid scar; gene therapy.

OS Canis familiaris.  
 PN WO200011033-A2.  
 XX 02-MAR-2000.  
 PD 25-AUG-1999; 99WO-US19329.  
 XX 25-AUG-1998; 98US-0097883.  
 PR 25-AUG-1998; 98US-0097883.  
 XX (LEXI-) LEXINGEN PHARM CORP.  
 PA (LEXI-) LEXINGEN PHARM CORP.  
 PI Lo K, Li Y, Gillies SD;  
 XX WPI; 2000-237616/20.  
 DR N-PSDB; AAZ51309.

PT Novel fusion protein of angiotensin or endostatin and an immunoglobulin  
 PT FC region, useful for treating conditions mediated by angiogenesis,  
 PT such as rheumatoid arthritis, tumors and macular degeneration -  
 XX Example 8: Pages 59-60; 68pp; English.

CC The patent discloses a DNA molecule encoding a fusion protein comprising  
 CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
 CC inhibitor selected from angiotensin, endostatin, a plasminogen fragment  
 CC having angiotensin activity, a collagen XVIII fragment having endostatin  
 CC activity, or combinations of them. The fusion protein (immunofusin) is

CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
 CC by angiogenesis. Conditions that may be treated include solid tumours,  
 CC blood born tumours, tumour metastasis, benign tumours including  
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
 CC plaque neovascularisation, telangiectasia, haemophilic joints,  
 CC angiodiroma, wound granulation, and excessive or abnormal stimulation of  
 CC endothelial cells, intestinal cells, atherosclerosis, scleroderma and  
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
 CC in gene therapy. The present sequence is a canine  
 CC endostatin used in the construction of immunofusin containing canine  
 CC immunoglobulin Fc fragment.

SQ Sequence 184 AA;  
 Query Match 86.9%; Score 841; DB 21; Length 184;  
 Best Local Similarity 84.2%; Pred. No. 6e-93;  
 Matches 155; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 HTHQDFOPVHLVNLTPLSGSGMRIGADPCQARAVGLSGFRFLSSRLQDLKST 60  
 Db 1 hthqdfqpyhlvnltpslsgsmrgirgadfqrtnar-vqlsgrlraflssrlqdlkyl 60  
 QY 61 VRADRGSPVIV-NLKDEVLSFSDSLFSGSQGLQPCARIFSPGDRVLRHPAPQKSV 120  
 Db 61 vrradrgspvivrnlkdevlfspswslfsgsqglqpcarifsfqdrvrlrhpapqrsy 120  
 QY 121 HGSOPSGRRLMESYCEWRTETTGATGQASSLSGRLLBOKRASCCHNSYIVLCIENSFM 180  
 Db 121 hgsdpsgrllmesycewrtetgtgassllsgrlllegraaschdsyivlcienfsm 180  
 QY 181 SFSK 184  
 Db 181 tsfsr 184

RESULT 14  
 AAB28399  
 ID AAB28399 standard; Protein; 182 AA.  
 AC AAB28399;

DT 19-FEB-2001 (first entry)  
 XX Human endostatin.  
 DE Human endostatin.

KW Human: endostatin; cytostatic; antiproliferative;  
 KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
 KW cancer; vascularised solid tumour.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN WO200064946-A2.  
 XX 02-NOV-2000.  
 PD 28-APR-2000; 2000WO-US11367.  
 XX 28-APR-1999; 99US-0131432.  
 PR 28-APR-1999; 99US-0131432.  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PI Thorpe PE, Brekken RA;  
 XX WPI; 2000-687317/67.

PT Immunogenic composition for the treatment and diagnosis of cancer  
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody  
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -

XX Example 10: Page 291-292: 298pp: English.

PS The present invention relates to anti-vascular Endothelial Growth Factor

CC (VEGF) antibodies that bind to the same epitope as the monoclonal

CC antibody AACC PTA 1595 and which significantly inhibit VEGF binding to

CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF

CC receptor VEGFR1. The present sequence is human endostatin. Endostatin

CC may be conjugated onto the anti-VEGF antibodies of the present invention.

CC The anti-VEGF antibodies of the present invention are useful for the

CC treatment and diagnosis of cancer, especially vascularised solid tumours.

CC

XX Sequence 182 AA:

QY Query Match 86.8%; Score 840; DB 21; Length 182;

Best Local Similarity 85.6%; Pred. No. 7.7e-93;

Matches 155: Conservative 15; Mismatches 11; Indels 0; Gaps 0;

DB 1 HTHQDFQPVHLVATLNTPLSGMRGIRGADFOCFQOARAVGLSGFRFAFLSRLODLYSI 60

1 hshrdtqpvhlvlnlspisgmrgrlgadfgcfqaravglagtrflsrldlysi 60

QY 61 VRRADGSPVPIVNLKDEVLSFSGSQGLQPGARIFSGRDVLRHPAMPQKSVW 120

61 vrradtraavpivnlkdelifsgseglkpgariffsgdkvrlrhpmpqksw 120

DB 121 HGSDPSGRRLMESYCEWTWRTTGTGATGQASSLLSGRLLEQKASCHNSYIVLCIENSFWT 180

121 hgsdpsgrrlmesyctwrttggatgqassllsgrrlleqkaschneyivlcienfwt 180

QY 181 S 181

DB 181 a 181

RESULT 15

ID AAY94323 standard; Protein: 182 AA.

AC AAY94323:

DT 11-AUG-2000 (first entry)

XX Human endostatin protein.

DE Human endostatin protein.

XX Human: endothelial cell proliferation inhibitor; collagen XVIII;

KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;

KW vasotropic; dermatological; ophthalmological; vulnary;

KW antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;

KW ocular angiogenic disease; atherosclerosis; scleroderma;

KW myocardial angiogenesis; telangiectasia; angiodysplasia;

KW wound granulation.

XX Homo sapiens.

OS WO200026368-A2.

PN 11-MAY-2000.

PD 01-NOV-1999: 99WO-US25605.

XX 30-OCT-1998: 98US-0106343.

PR 20-MAY-1999: 99US-0315689.

XX (CHIL-) CHILDRENS MEDICAL CENT.

PA O'Reilly MS, Folkman MJ;

XX WPI: 2000-365617/31.

DR N-PSDB, AAA27004.

XX Novel endostatin capable of inhibiting endothelial cell proliferation

PT and angiogenesis, useful for treating angiogenesis-dependent cancers

PT and as birth control agents

PS Claim 2: Page 38: 68pp: English.

XX The present sequence is an endostatin protein which is the carboxy

CC terminal protein of human collagen XVIII. Recombinant mouse endostatin

CC (20 mg/kg) was administered subcutaneously to mice implanted with Lewis

CC lung carcinomas. There was tumour mass regression non-detectable levels

CC after 12 days of therapy due to the angiogenesis inhibitory activity of

CC endostatin. Thus the protein is useful for treatment of angiogenesis-

CC dependent cancers. The polynucleotide and diagnosis of tumours, ocular

CC endostatin are useful for treating or preventing diseases related to

CC angiogenic diseases, Osler-Webber syndrome, myocardial angiodysplasia,

CC plaque neovascularization, telangiectasia, haemophilic joints,

CC angiodysplasia and wound stimulation of endothelial cells e.g. intestinal

CC adhesions, atherosclerosis, scleroderma. The protein may also be useful

CC as a birth control agent by reducing or preventing uterine

CC vascularisation. The gene for endostatin may be isolated from cells or

CC tissue that express high levels of endostatin, eg. tumour cells, by

CC generating cDNA from mRNA using reverse transcriptase and then amplifying

CC the DNA sequence.

XX Sequence 182 AA:

QY Query Match 86.8%; Score 840; DB 21; Length 182;

Best Local Similarity 85.6%; Pred. No. 7.7e-93;

Matches 155: Conservative 15; Mismatches 11; Indels 0; Gaps 0;

DB 1 HTHQDFQPVHLVATLNTPLSGMRGIRGADFOCFQOARAVGLSGFRFAFLSRLODLYSI 60

1 hshrdtqpvhlvlnlspisgmrgrlgadfgcfqaravglagtrflsrldlysi 60

QY 61 VRRADGSPVPIVNLKDEVLSFSGSQGLQPGARIFSGRDVLRHPAMPQKSVW 120

61 vrradtraavpivnlkdelifsgseglkpgariffsgdkvrlrhpmpqksw 120

DB 121 HGSDPSGRRLMESYCEWTWRTTGTGATGQASSLLSGRLLEQKASCHNSYIVLCIENSFWT 180

121 hgsdpsgrrlmesyctwrttggatgqassllsgrrlleqkaschneyivlcienfwt 180

QY 181 S 181

DB 181 a 181

Search completed: August 28, 2001, 12:51:12

Job time: 127 sec



